Journal of Nanoscience and Nanotechnology
Vol. 5, 1759–1767, 2005

REVIEW

Rational Design of Thermally Stable Proteins: Relevance to Bionanotechnology

V. Renugopalakrishnan,[1,*] R. Garduño-Juárez,[2] G. Narasimhan,[3]
C. S. Verma,[4] X. Wei,[3] and Pingzuo Li[1,5]

[1] *Children's Hospital, Harvard Medical School, Harvard University, Boston 02115, USA*
[2] *Centro de Ciencias Físicas, Universidad Nacional Autónoma de México, 62210 Cuernavaca, México*
[3] *Bioinformatics Research Group (BioRG), School of Computing and Information Sciences, Florida International University, University Park, Miami, FL 33199, USA*
[4] *Bioinformatics Institute, 30 Biopolis Street, 07-01 Matrix, Singapore-138671, Singapore*
[5] *Shanghai Research Center for Biotechnology, Chinese Academy of Sciences, Shanghai 200233, P. R. China*

Design of thermally stable proteins is spurred by their applications in bionanotechnology. There are three major issues governing this: first, the upper limit on the temperature at which proteins remain physiologically active and are available for technological applications (answers may emerge from the discovery of new, natural hyperthermophilic enzymes that are active above 125 °C or from the selection of mutants of hyperthermophilic enzymes that are more stable); second, the use of hyperthermophilic enzymes as molecular templates to design highly stable enzymes that have high activity at low temperatures; third, the link between rigidity and flexibility to thermostability and activity, respectively. We review progress in these areas.

Keywords: Thermally Stable Protein, Rational Design, Protein Thermal Stability, Bacteriorhodopsin, Rubredoxin, Ferredoxin.

CONTENTS

1. INTRODUCTION

Bionanotechnology attempts to harness various functions of biological macromolecules and integrates them with engineering for technological applications. It is based on a bottom-up approach and encompasses structural biology, biomacromolecular engineering, materials science, and engineering, and thus extends the horizon of materials science. The structural biology of proteins has been developed over the last 70 or so years. Of these seven decades, the last three have seen a transformation of our static perspective of proteins to dynamical entities. This has laid the basis for rational engineering and design of proteins with enhanced and novel properties. Biotechnology has matured during the same period of time and has facilitated the design of proteins. Protein self-assembly in suitable media offers unique advantages in the fabrication of protein-based nanodevices, and circumvents cost-prohibitive processes in the manufacture of bioelectronic devices. Bionanotechnological applications of proteins require thermally stable proteins.

Thermal stabilization of proteins[1–7] has been a focus of intense research since the discovery of thermophilic and hyperthermophilic organisms.[8,9] The stereochemical basis of heat stability in bacterial ferredoxins and hemoglobin were discussed long ago by Perutz and Raidt,[10] and Perutz.[11] Despite advances in our understanding of thermophilic and hyperthermophilic organisms and their

*Author to whom correspondence should be addressed.

 1533-4880/2005/5/1759/009 doi:10.1166/jnn.2005.441

REVIEW

V. Renugopalakrishnan is currently Professor, Harvard Medical School, Children's Hospital, Boston, Massachusetts. He will be directing the Bionanotechnology research group. He obtained his B.Sc. from the Madras University and Ph.D. in biophysics from Columbia/State University of New York, Buffalo, New York. He is a biophysicist by training and his research interests have been focused on protein structure, dynamics, engineering for over quarter of a century. In the last five years his laboratory has been very active in technological applications of proteins in information storage and more recently on protein-based biosensors, hydrogen fuel cells. He is the author of 200 publications, 7 patents and member of numerous academies. He is an Editor-in-Chief of the Journal of Bionanoscience published by American Scientific Publishers.

R. Garduño-Juárez is a Professor of biophysics at the Universidad Nacional Autónoma de México. He is a Biochemical Engineer with a Ph.D. in biophysics from the State University of New York at Buffalo. He is a member of the Mexican Academy of Sciences. He was a recipient of a fellowship from the National Research Council and the National Academy of Sciences for post-doctoral research at NASA Ames Laboratories. He has been a guest scientist at Roswell Park Memorial Institute in Buffalo NY, at Cornell University and at the Universidad Autónoma de Barcelona, Spain. He has published extensively on heuristic methods applied to the protein folding problem and uses the tools of computational chemistry to solve problems in chiral selectivity by cyclodextrins.

G. Narasimhan obtained his B.Tech. from the Indian Institute of Technology, Mumbai, India, in Electrical Engineering, and his Ph.D. from the University of Wisconsin, Madison, USA. He was part of the Computer Science faculty in the Mathematical Sciences Department at the University of Memphis from 1989 to 2001. He is currently a Professor in the School of Computer Science at Florida International University, Miami, and heads the Bioinformatics Research Group (BioRG). His research interests also include design and analysis of algorithms and bioinformatics.

C. S. Verma obtained his D.Phil. from the University of York in the United Kingdom in 1990—having worked on kinetic and structural models of cisplatin binding to DNA. He then joined the laboratory of Structural Biology headed by Prof. G. G. Dodosn FRS on a Wellcome fellowship and remained there until 2003 when he moved to take up a PI position at the Bioinformatics Institute in Singapore, where he heads a group in computational structural biology and cheminformatics. He holds a joint appointment as an adjunct faculty in the Department of Biological Sciences at the National University of Singapore. He has been using computational techniques to understand mechanisms underpinning functions in proteins.

X. Wei obtained her B.S. degree in Microbiology from Sichuan University in China. She received her M.S. in Computer Science from Florida International University, and did her thesis work under the guidance of Prof. Giri Narasimhan. She is currently a Ph.D. student in the Department of Computer Science at Tufts University, pursuing her research interests in Bioinformatics.

Pingzuo Li obtained his M.S. in Fermentation Engineering, Tianjin Institute of Light Industry, Tianjin, China and his Ph.D. in Biochemical Engineering, Wuxi University of Light Industry, Wuxi, China. He is currently an Associate Professor at the Shanghai Research Center for Biotechnology, Chinese Academy of Sciences, Shanghai, China. He is expecting to join Dr. Renugopalakrishnan's group at Harvard in 2006.

protein constituents, the molecular determinants of thermal stability of proteins have remained unclear. Recent advances in structure-stability analyses[12–18] coupled with mining of bioinformatics databases of mesophilic, thermophilic and hyperthermophilic proteins have identified several features as major contributors to thermal stability.[19–23] These in turn have helped in establishing some critically important principles for the design of

thermally stable proteins needed in the fabrication of biomolecular devices in bionanotechnology.[24–26]

Earlier investigations have proceeded along two major avenues. The first approach was to compare the structures and sequences of homologous proteins from hyperthermophiles, thermophiles, and mesophiles,[27–29] which resulted in the identification of various differentiated features related to the thermal stability, such as an increased number of salt bridges, better hydrogen bonding, higher internal packing, stronger inter-subunit associations, etc.[30–33] The second major approach involved protein engineering based on the central doctrine that protein thermostability is a direct consequence of its sequence.[34–40]

Mutations causing enhancement of thermal stability of a protein either increase the thermodynamic stability of a protein, ΔG, between unfolded and folded states in the denaturation process or they decrease the rate of unfolding by increasing, $\Delta G^{\#}$, between the folded state and the transition states.[41] Commonly adopted strategies include decreasing the ΔS of the unfolded state by: introducing additional disulfide bridges, making X $\rightarrow$ Pro mutations, increasing the α-helical propensity through Gly $\rightarrow$ Ala mutations, stabilization of α-helix macrodipoles[42] by imposing electrostatic interactions between charged surface residues, introducing additional salt bridges or salt bridge networks, or optimizing the electrostatic potential surfaces.[43]

2. FACTORS THAT FACILITATE PROTEIN THERMO STABILITY

Hydrophobic Effect: The hydrophobic effect is considered to be the major driving force behind protein folding,[44–46] leading to a collapsed structure from which the native structure is defined by the interplay of several intermolecular forces, e.g., H bonds, ion pairs, and Van der Waals. The following are two critical observations about the thermal stability imparted by the conserved protein core of mesophilic and hyperthermophilic homologues: (i) hydrophobic interactions and residues involved in secondary structures in the core are better conserved than on the surface, and (ii) numerous stabilizing substitutions are found in solvent-exposed areas. The high level of similarity encountered in the core of mesophilic and hyperthermophilic protein homologues suggests that both proteins are equally optimally packed. Stabilizing interactions in hyperthermophilic proteins are often found in the less conserved areas of the protein. However, it is becoming clear there are several factors that contribute to the remarkable stability of hyperthermophilic proteins.[47,48]

Electrostatics: The role of electrostatics in protein stability has been an enigmatic issue. While some report a degree of correlation between increased stability and the number of salt bridges between thermophilic and mesophilic proteins,[49–52] an earlier comprehensive analysis[53] had found that although in general the hyperthermophilic proteins had more favorable electrostatic interactions, there was no direct correlation with the number of ionizable residues, ion pairs, or ion-pair networks. They suggested that optimized electrostatics resulting from optimal location of amino acids would give rise to the enhanced stabilities. Indeed this is a consensus reached by an increasing number of recent studies.[44,54,55] Torrez et al. computed the stabilities of several folded proteins and found that a combination of factors, notably the improvement of surface electrostatics, with origins in no specific ion pair, was responsible for enhanced thermostability in cold shock proteins, such as RNase T1 and CheY.[49] In contrast, Makhatadse et al.[56] optimized the electrostatics and stabilized proteins by decreasing the enthalpy–entropy of unfolding. This work, in line with the philosophy of Xiao and Honig,[53] lead to the successfully engineering of the surface charge residues of a thermophile on to the core residues of a mesophile, enhancing its thermal stability.

Disulfide Bridges: The influence of disulfide bridges on protein thermal stability[57] and folding has been investigated experimentally using phage T4 lysozyme,[58] hen lysozyme,[59] barnase,[60] amylase inhibitor tendamistat,[61] azurin mutant,[62] and theoretically[63] by the removal of natural –SS– as well as the insertion of novel –SS– bridges in a protein. It is clear that the removal of a natural –SS– bond is usually accompanied by a decrease of protein thermal stability due to higher conformational entropy and hence stabilization of the unfolded state.

Aromatic Interactions: Aromatic–aromatic interactions (aromatic pairs) are important contributors to the thermostability of proteins. In a recent study[64] a graph spectral method was used to identify aromatic clusters in a dataset of 24 protein families. The study concluded that the presence of additional aromatic clusters enlarged the aromatic networks in 17 different thermophilic protein families, which is in marked contrast to their mesophilic counterparts. The aromatic clusters form relatively rigid regions of the surface and often the additional aromatic cluster is located close to the active site of the thermophilic enzyme. The residues in the additional aromatic clusters are preferentially occupied by Leu, Ser or Ile in the mesophilic homologue. The following characteristics of aromatic pairs were extracted from analysis of 272 aromatic pairs in 34 high-resolution structures of mesophilic proteins:[65] in two-thirds of the pairs, the interacting rings are close to perpendicular; most are involved in a network; most link distinct secondary structural elements; most are energetically favorable (80% have potential energies between 0 and −2 kcal/mol); and most take place between buried or partially buried residues. For example, thermitase,[66] the serine proteinase produced by *Thermoactinomyces vulgaris*, contains 16 aromatic residues involved in aromatic pairs; the mesophilic homologue *Bacillus amyloliquefaciens* subtilisin BPN contains only 6 aromatic pairs.

Single/Multiple Mutations: Recently, there have been several experimental reports of proteins displaying appreciable stability gains through mutation of one or two amino acid residues.[67] Perl et al.[68] have reported the mutation of Glu residues at positions 3 and 66 to Arg and Leu caused a significant change in their thermal stability. The results suggest that elimination of like-charge repulsions and creation of opposite-charge attractions on the protein surface is an efficient method to confer thermostability to a mesophilic protein. In contrast, Lehmann et al.[21,69] found that in fungal phytases, which belong to the family of histidine acid phosphatases,[70,71] thermostabilization was achieved as a result of a combination of slight improvements from multiple mutations rather than the effect of a single or just a few dominant mutations that may have been introduced by chance.

3. STRATEGIES FOR ENGINEERING PROTEIN THERMOSTABILITY

Random Mutagenesis: There are three widely used strategies for random mutagenesis. The first is oligonucleotide-directed mutagenesis. Oligonucleotides can be synthesized to contain predetermined nucleotide alterations to obtain specific codons. The second one is to use the error-prone polymerase chain reaction (PCR) to replicate the target gene. The third method is DNA shuffling, whereby a set of similar genes are spliced into pieces and then regenerated using a thermostable DNA polymerase[72] allowing multiple mutations to recombine *in vitro*. To efficiently screen the mutants, an *in vitro* phage display technology was introduced[73] whereby a library of proteins is fused to a gene that codes for one of the phage coat proteins and the mutants can be screened for binding to a target. In this way a small number of active proteins can be separated from millions of inactive variants.[74]

Rational Site-Directed Mutagenesis: Rational site-directed mutagenesis relies on the principle of optimizing a chosen physical property by analyzing how that property depends on essential amino acids at specific locations in the protein by substituting them by suitable candidates found at the same site in similar protein families.

Directed Evolution: The Directed Evolution approach utilizes repeated rounds of random mutagenesis, starting with a given parent gene of interest. After each round, the best mutants are selected and used as parent sequences in the next round of random mutagenesis. It is a laborious, time-consuming process.

Consensus Sequences: The Consensus sequence approach relies on sequences derived by alignment of homologous proteins manifesting inherent thermostability.[75,76] There have been numerous attempts to combine these approaches to study thermal stability. In this review, we discuss various approaches to achieve enhanced thermal stability.

4. RATIONAL DESIGN OF THERMALLY STABLE PROTEINS

4.1. Thermodynamic Force Field for Proteins

The design of a thermally stable protein must include the following criteria:

(a) The side chains of the amino acid substituted to enhance T_m must be sterically compatible within the backbone structure.
(b) The mutated sequence must have a ΔC_p value that can be achieved by burying more apolar surface area or by exposing more polar surface area in the folded protein.

The hydrophobic core of the predicted protein must have low conformational entropy of folding to ensure a unique and rigid internal architecture.

4.2. Computational Methods Identifying Mutations that Enhance Protein Thermostability

The program "Rational site-directed mutagenesis selection algorithm" (RSDMSA) (Renugopalakrishnan, US Patent Disclosures), which is an improvement over an earlier program (CORE) developed by Jiang et al.,[77] predicts protein hydrophobic core sequences that can fold into a target backbone structure. Basically RSDMSA rejects or eliminates unfavorable mutations starting from an input backbone.

A typical prediction run in RSDMSA starts by randomizing the sequence of hydrophobic core of a selected protein to ensure that no bias is introduced at the start of a run. After this, a simulated annealing run driven by the Metropolis algorithm is initiated. Following the last step, a single mutation of a core residue chosen based on the specific property and nature of the protein is allowed. Residues are allowed to mutate to Ala, Ile, Met, Phe, Tyr, Trp, or Val. RSDMSA then initiates a nested simulated annealing run to determine the best rotamer configuration of all core residues. The simulated annealing run reveals the number of unfavorable van der Waals interactions by calculating the number of hard sphere bumps. If the number of bumps for the simulated annealing run is greater than zero, the sequence is rejected. If the number of bumps is zero, a second simulated annealing on the same sequence is initiated followed by a low temperature Monte Carlo run which yields two parameters:

(i) mobility of each amino acid in the form of conformational entropy which is averaged to obtain global conformational entropy for the whole protein ($\Delta S_{conf.}$) as shown below

$$\Delta S = \sum_{1}^{n} \Delta S_{\text{conf}}^{\text{residue}}$$

(ii) heat capacity (ΔC_P) of the hydrophobic core of the protein calculated based on heat capacity of individual residues.[2]

These two parameters, plus the number of bumps, are used to calculate the "scores" for this particular sequence. This score then drives the main sequence simulated annealing run. After 10 sequences are sampled at each temperature, the Metropolis temperature is gradually decreased during the simulated annealing process. This gradual decrease in the Metropolis temperature ensures that local minima are avoided by slowly lowering the probability that sequences are accepted with scores higher than the previously accepted sequence. The simulated annealing run is automatically terminated when the number of accepted sequences is consistently zero. The temperature at which this occurs, or an arbitrarily low temperature, T, is used for a final sequence Monte Carlo run initiated with the sequence determined from the simulated annealing run. This Monte Carlo run is conducted to sample the sequence space around the simulated annealing sequence. A large value of T allows the program to sample from sequences with wider range of scores around the score for the simulated annealing sequence. Typically these runs generate a family of 100–1000 sequences of proteins with good thermal stability depending on the value of T and the number of core residues mutated.

It is worth mentioning the recent application of the Rosetta Design program in the computational thermostabilization of the yeast cytosine deaminase,[1] by which a 10 °C increase in the apparent T_m and a 30-fold increase in half life at 50 °C was obtained with no reduction in its catalytic efficiency.

5. THERMOSTABILITY STUDIES IN SPECIFIC PROTEIN FAMILIES

Bacteriorhodopsin (bR): Nanoelectronic devices based on bR require thermally robust mutants of bR.[78] Therefore increasing the thermostability of bR has been an area of intense research. One method of choice has been rational site-directed mutagenesis of bR. Advancements in thermostable vectors,[79] antibiotic resistance genes, and the genetic characterization of extreme thermophiles have prompted the development of *in vivo* thermoselection systems to optimize mesophilic proteins bR for device applications.

Thermus thermophilus has been a useful *in vivo* screening platform for bR mutants; a versatile, heat-stable expression vector is required. Moreno et al.[79] constructed a bifunctional vector system (pMKE1) capable of expression in both extremely thermophilic (*T. thermophilus*) and mesophilic (*E. coli*) microorganisms. Mutants that retain structural stability at elevated temperatures have been used as starting points for additional rounds of mutagenesis and thermoselection. Several iterations of thermoselection may be required before a bR variant with adequate thermostability can be used for device applications. In a recent study to optimize the heat capacity of bR,[2] eight residues were selected in bR: D85, W86, L93, D96, D115, W182, W189, D212. Differential Scanning Calorimetric studies of bR mutants designed and expressed manifest $T_m \sim 192$ °C.

Rubredoxin: Rubredoxin (Rd) from the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, an organism that grows optimally at 100 °C and above, has been used as a test-bed for thermal stabilization of proteins. The thermal denaturation of rubredoxin has been previously shown to exhibit unusual kinetics[80] characterized by nearly constant rate of change and a reaction time independent of the starting amount of the protein. Hiller et al.[81] have discussed the stability and dynamics of rubredoxin with melting temperatures close to 200 °C.

Rubredoxins are probably the simplest members of the large ubiquitous family of redox metalloenzymes and consist of a relatively short polypeptide chain (~53 AA). Their structural features are fairly well known because the primary structures of 15 bacterial rubredoxins, 10 X-ray crystal structures, and an NMR solution structure of Zn substituted Pf RD have been reported to date.[82] A noteworthy feature of all Rd structures is a triple-stranded antiparallel β-sheet stretch consisting of 14 or 15 residues of the N-terminal and the last 5 or 6 residues at the C-terminal. Among the conserved residues are the four Cys residues from the active site and the five aromatic residues that constitute the hydrophobic core of the proteins. The aromatic cluster in the hydrophobic core contributes significantly to its thermal robustness.

The RSDMSA program was applied to rubredoxin to select mutants with high thermal stability. Optimization of rubredoxin thermodynamics was achieved by increasing its total heat capacity, and consequently its T_m to 200 °C (Renugopalakrishnan et al., in preparation). The underlying rationale for the selection of residues A9, A13, A31, A42, A43, and A51 for site directed mutagenesis rests on the hypothesis that the thermostability is conferred by interaction between β-sheet segments 1 and 2[83] (Renugopalakrishnan et al., unpublished). Three designed mutants of rubredoxin, Figure 1, (A9F, A13W, A43I; A9G, A13D, A31E, A42G, A51L; A9Y, A31L, A42F, A51I) exhibit a T_m in the vicinity of 200 °C. Detailed DSC studies have been performed and will be reported in the literature (Renugopalakrishnan et al., in preparation).

Ferredoxins: Ferredoxins are a group of small, iron–sulfur electron transport proteins that show no enzymatic function, unlike the way complex iron–sulfur proteins behave.[84] They are involved in crucial metabolic processes such as photosynthesis, oxidative phosphorylation, and nitrogen fixation. Three types of structurally characterized iron-sulfur centers, distinguished by the number of iron and sulfur atoms, have been found in ferredoxins and are called: [2Fe–2S] (found mainly in plant

REVIEW

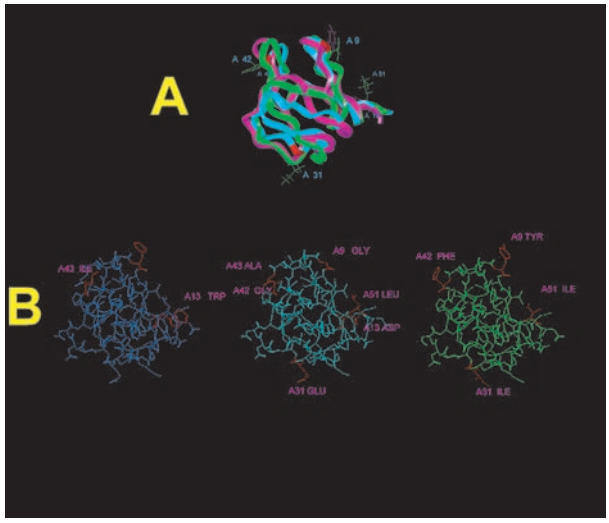

Fig. 1. 3D structure of wild type rubredoxin from Pyrococcus furiosus residues mutated are shown marked.

and animal ferredoxins), [3Fe–4S], and [4Fe–4S] (found mainly in prokaryotes). The family of ferredoxins includes ferredoxins with one [2Fe–2S], ferredoxins with one or more [4Fe–4S], ferredoxins with a single [3Fe–4S], and ferredoxins with one [3Fe–4S] and one [4Fe–4S] cluster. Devanathan et al. had suggested that increased hydrogen binding was the reason for the increased stability of thermophilic ferredoxins.[85]

The PROSITE signature, PS00198, for the [4Fe–4S] type ferredoxins is: **C–x–{P}–C–x(2)–C–{CP}–x(2)–C–[PEG]**, and is reported to have a precision and recall of about 96% from Swiss-PROT database. Note that curly braces indicate residues that are disallowed. This pattern also has a precision and recall of 100% and 92% for proteins from the ASTRAL95 database[86] (set of proteins for which structures are known and which have at most 95% similarity to each other). An improved signature,[47] SSP99243, given by: **C–[WTLIMN]–x(1)–C–x(2)–C–x(3)–[AC]** was developed using the SSP method.[87] SSP99243 has a precision and recall of 100% and 99% from the ASTRAL95 database. Most ferredoxins contain two or more occurrences of the above signatures. Normally, one [4Fe–4S] cluster is ligated by the first three cysteines in one signature (highly conserved) and the fourth cysteine from a more remote part of the peptide chain.[84] Another signature SSP00298, given by **[ADQSV]–C–[AEGIKV]–[AENPRS]–[AEILV]–x(4, 5)–[IFLY]–x(19, 31)–AC–x(3)–C–P**, accurately describes the SCOP families D.58.1.1 through D.58.1.4, while SSP99243 additionally describes SCOP family A.1.2.1.[47]

Further sequence analysis showed that in terms of their amino acid compositions, Asn and Glu are decreased in (hyper)thermophiles, while Arg is increased. Hyperthermophilc ferredoxins are smaller than their thermophilic counterparts, which in turn are smaller than their mesophilic counterparts.

Table I. Consensus sequences for proteins that strive from hot to cold environments.

Extremophile	Stabilizing factors	Ref.
Hyper thermophiles	Reduced in Gln, increased in Glu	[89]
	Salt bridges rigidify the protein structure	[91]
	Large number of Ile	[92]
	Decrease of Gly in α-helices	[100]
Thermophiles	Reduced in Gln, Gly $\rightarrow$ Ala, Lys $\rightarrow$ Arg	[89]
	Arg and Tyr are more frequent; Cys and Ser are less frequent. Longer α-helices, avoid Pro in α-helices	[59]
	Polar residues form extra hydrogen bonds with each subunit	[90]
	Charged residues form extra ionic interactions between dimer–dimer interface	[92, 93]
	Higher number of Ala and aromatic residues	[100]
	Decrease of β branched residues (Val, Ile, Thr)	[92]
	Larger Thr:Ser ratio in β-strands	[100]
	Increase of charged residues on the protein surface, especially Arg. Lower frequency of Asn	
Psychrophiles	Arg and Glu are replaced in exposed sites of α-helices by Lys and Ala, in the direction of 'hot' to 'cold' enzymes	[94]
	Strictly conserved residues near the metal binding sites	[95]
Consensus motifs or sequences		
Thermophiles	AXXXA found more often in α-helices of thermophiles than the GXXXG found in mesophiles	[96]
	X_2RAIYDVIGPX$_2$DIPAPDVYTNPQIMAWX around residues 50 to 79 on glutamate dehydrogenases	[97]
	135 amino-acid long consensus sequence for azurin and plastocyanine	[99]
Psychrophiles	331 amino-acid long consensus sequence for A_4-LDH's of Antartic fishes	[98]
	10 non-conservative differences in the amino-acid sequence. Of these two Gly are found, 10 residues apart, in an important loop in one side of the catalytic vacuole. A Pro $\rightarrow$ Ala at the N-terminal of a helix bordering the catalytic vacuole	
	Alkaline protease binds eight Ca^{2+} important for stabilization. Two Ca^{2+} bound to a region between the N-terminal domain and the C-terminal domain. Five calcium ions (Ca3–Ca7) located near a β roll with consensus sequence GGXGXDXUX (U is a large hydrophobic residue). The eighth Ca^{2+} is located within the consensus sequence SGDAHAD (residues 448–454). The consensus sequence HEXXH(X)$_4$SH (residues 176–186) involves three His in the Zn^{2+} coordination	[95]

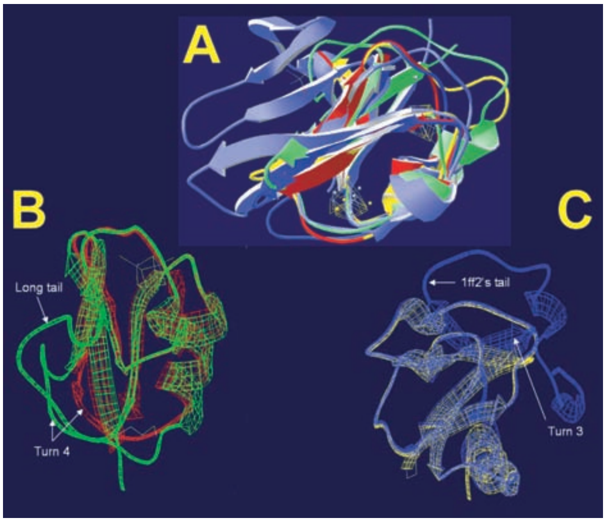

Fig. 2. Ferredoxin structures: A. Figure shows how the Fe–S binding site is structurally aligned for a number of different ferredoxins. B. Structure alignment of hyperthermophilic ferredoxin 1vjw (red) and thermophilic ferredoxin 1iqz (green) is shown. Thermophilic 1iqz has a longer turn 4 and a longer tail than hyperthermophilic protein 1vjw. C. The structure alignment of thermophilic Ferredoxin 1h98 (yellow) and mesophilic Ferredoxin 1ff2 (blue) is shown. While there is small misalignment in turn 3, mesophilic 1ff2 has a longer tail.

Structural analysis of ferredoxins (see Fig. 2) shows that they possess two α-helices, two antiparallel β-sheets and four turns. Turn 3 and turn 4 are most flexible. The β-turns and tails tend to be smaller in more thermostable ferredoxins.

Consensus Sequences from Other Protein Families: Despite the many protein sequences that are deposited every year in the protein data banks, only a few of these report the corresponding X-ray structures. Since the 3D structures of thermally stable proteins isolated from many extremophiles are very similar, in many cases only qualitative data can be obtained in order to explain their thermostability at either cold or high temperatures. The qualitative information include data such as the relative ratios of some amino acids between (hyper)thermophilic proteins and similar mesophilic proteins, and some differences on structural motifs. In other cases researchers have used the nucleic acid sequences from the mapped genes of several organisms that dwell in the same environment. From this information several amino acid consensus sequences have been reported for psychrophiles, hyperthermophiles, thermophiles, and mesophiles. A short list of the consensus sequences reported in the last six years is provided in Table I.

6. CONCLUSIONS

Protein thermodynamic stability, ΔG, is determined by a multitude of intra- and inter-molecular interactions, which in turn depend on the primary amino acid sequence of the protein. The primary sequence determines all secondary and tertiary structural features. Hence it is the sequence that determines protein thermodynamic stability. At a structural level this will consist of essential stability determined by the packing of the secondary structural elements which will be complemented by the interactions of and between sidechains. Berezovsky and Shakhnovich[88] recently concluded that most strategies for rational design of thermally stable proteins are of two types: (i) non-specific, structure-based strategies, with contributions from different stabilizing interactions, and (ii) specific, sequence-based strategies, using only dominating factors for adaptation to extreme conditions. The rapidly increasing number of fully sequenced genomes will be an invaluable help in deciphering which sequence variations among homologous proteins are related to stability and which ones are simply a result of evolution. Bionanotechnological applications demand a variey of thermally stable proteins in the design of bionanosensors, bionanomemory, and biofuel cells.

Acknowledgments: Renugopalakrishnan expresses his thanks to National Science Foundation, Harvard Medical School, Wallace H. Coulter Foundation, and Office of Naval Research (ONR) for supporting this project. Research of GN was supported in part by NIH Grant P01 DA15027-01. We thank Tom Milledge for help with the SSP program. BII is an ASTAR Institute.

References and Notes

1. *Bionanotechnology*: *Proteins to Nanodevices*, edited by V. Renugopalakrishnan, R. V. Lewis, and P. K. Dhar, Springer, The Netherlands **(2005)**; A. Korkegian, M. E. Black, D. Baker, and B. L. Stoddard, *Science* 308, 857 **(2005)**.
2. V. Renugopalakrishnan, X. Wei, G. Narasimhan, S. L. Lakka, C. S. Verma, P. Li, and A. Anumanthan, in *Bionanotechnology: Proteins to Nanodevices*, edited by V. Renugopalakrishnan, R. V. Lewis, and P. K. Dhar, Springer, The Netherlands, in press.
3. H. K. Liang, C. M. Huang, M. T. Ko, and J. K. Hwang, *Proteins: Structure, Function, and Bioinformatics* 59, 58 **(2005)**.
4. B. S. Ibrahim and V. Pattabhi, *Biochem. Biophys. Res. Commun.* 325, 1082 **(2004)**.
5. S. Adinolfi, M. Nair, A. Politou, E. Bayer, S. Martin, P. Temussi, and A. Pastore, *Biochemistry* 43, 6511 **(2004)**.
6. A. Mozo-Villarias, J. Cedano, and E. Querol, *Protein Engineering* 16, 279 **(2003)**.
7. R. Das and M. Gerstein, *Functional and Integrative Genomics* 1, 76 **(2000)**.
8. R. Sterner and W. Liebl, *Crit. Rev. Biochem. Mol. Biol.* 36, 39 **(2001)**.
9. K. O. Stetter, *FEMS Microbiol. Rev.* 18, 149 **(1996)**.
10. M. F. Perutz and H. Raidt, *Nature* 255, 256 **(1975)**.
11. M. F. Perutz, *Differentiation* 13, 47 **(1979)**.
12. T. Dams, G. Auerbach, G. Bader, U. Jacob, T. Ploom, R. Huber, and R. Jaenicke, *J. Mol. Biol.* 297, 659 **(2000)**.
13. G. Auerbach, R. Ostendorp, L. Prade, I. Korndorfer, T. Dams, R. Huber, and R. Jaenicke, *Structure* 6, 769 **(1998)**.
14. Y. I. Chi, L. A. Martinez-Cruz, J. Jancarik, R. V. Swanson, D. E. Robertson, and S. H. Kim, *FEBS Lett.* 445, 375 **(1999)**.
15. R. J. Russell, J. M. Ferguson, D. W. Hough, M. J. Danson, and G. L. Taylor, *Biochemistry* 36, 9983 **(1997)**.

16. K. P. Hopfner, A. Eichinger, R. A. Engh, F. Laue, W. Ankenbauer, R. Huber, and B. Angerer, *Proc. Natl. Acad. Sci. USA* 96, 3600 **(1999)**.
17. T. H. Tahirov, H. Oki, T. Tsukihara, K. Ogasahara, K. Yutani, K. Ogata, Y. Izu, S. Tsunasawa, and I. Kato, *J. Mol. Biol.* 284, 101 **(1998)**.
18. M. M. Gromiha, M. Oobatake, and A. Sarai, *Biophys. Chem.* 82, 67 **(1999)**.
19. R. Jaenicke, *FASEB J.* 10, 84 **(1996)**.
20. C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala, *FASEB J.* 10, 75 **(1996)**.
21. M. Lehmann and M. Wyss, *Curr. Opin. Biotech.* 12, 371 **(2001)**.
22. C. Vieille, D. S. Burdette, and J. G. Zeikus, *Biotech. Annu. Rev.* 2, 1 **(1996)**.
23. J. Hollien and S. Marqusee, *Biochemistry* 38, 3831 **(1999)**.
24. C. S. Verma and V. Renugopalakrishnan, in *Modeling and Simulating Materials Nanoworld, Advances in Science and Technology*, edited by P. Vincenzini and F. Zerbetto, Techna Group, Srl. **(2004)**, Vol. 44, p. 321.
25. B. van den Burg, A. de Kreij, P. van der Veek, J. Mansfeld, and G. Venema, *Biotech. Appl. Biochem.* 30, 35 **(1999)**.
26. T. J. Graddis, R. L. Remmele, Jr., and J. T. McGrew, *Current Pharmaceutical Biotechnology* 3, 285 **(2002)**.
27. G. J. Davies, S. J. Gamblin, J. A. Littlechild, and H. C. Watson, *Proteins* 15, 283 **(1993)**.
28. D. Maes, J. P. Zeelen, N. Thanki, N. Beaucamp, M. Alvarez, M. H. Thi, J. Backmann, J. A. Martial, L. Wyns, and R. Jaenicke, *Proteins* 37, 441 **(1999)**.
29. Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi, *DNA Research* 5, 55 **(1998)**.
30. H. R. Bosshard, D. N. Marti, and I. Jelesarov, *J. Mol. Recogn.* 17, 1 **(2004)**.
31. S. Kumar and R. Nussinov, *Chem. Bio. Chem.* 5, 280 **(2004)**.
32. Z. S. Hendsch and B. Tidor, *Protein Sci.* 3, 211 **(1994)**.
33. C. D. Waldburger, J. F. Schildbach, and R. T. Sauer, *Nat. Struct. Biol.* 2, 122 **(1995)**.
34. G. Chakshusmathi, K. Mondal, G. S. Lakshmi, G. Singh, A. Roy, R. B. Ch, S. Madhusudhanan, and R. Varadarajan, *Proc. Natl. Acad. Sci. USA* 101, 7925 **(2004)**.
35. P. Argos, M. G. Rossman, U. M. Grau, H. Zuber, G. Frank, and J. D. Tratschin, *Biochemistry* 18, 5698 **(1979)**.
36. R. Varadarajan, H. A. Nagarajaram, and C. Ramakrishnan, *Proc. Natl. Acad. Sci. USA* 93, 13908 **(1996)**.
37. M. Fried, *Proc. Natl. Acad. Sci. USA* 53, 486 **(1965)**.
38. D. P. Goldenberg, D. H. Smith, and J. King, *Proc. Natl. Acad. Sci. USA* 80, 7060 **(1983)**.
39. B. Schuler and R. Seckler, *J. Mol. Biol.* 281, 227 **(1998)**.
40. C. Aubert, F. Guerlesquin, P. Bianco, G. Leroy, P. Tron, K. Stetter, and M. Bruschi, *Biochemistry* 40, 13690 **(2001)**.
41. B. Steipe, *Curr. Top. Microbiol. Immuno.* 243, 55 **(1999)**.
42. C. Cohen and D. A. D. Parry, *Proteins* 7, 1 **(1990)**.
43. R. R. Naik, S. M. Kirkpatrick, and M. O. Stone, *Biosens. Bioelectron.* 16, 1051 **(2001)**.
44. K. A. Dill, *Biochemistry* 29, 7133 **(1990)**.
45. L. Lins and R. Brasseur, *FASEB J.* 9, 535 **(1995)**.
46. K. Saraboji, M. Gromiha, and M. N. Ponnuswamy, *Int. J. Biol. Macromol.* 35, 211 **(2005)**.
47. X. Wei, T. Milledge, V. Renugopalakrishnan, and G. Narasimhan, **(2005)** Unpublished results.
48. P. Shcherbatov, J. S. Anderson, G. Hernandez, and D. LeMaster, Abstracts, 31st Northeast Regional Meeting of the American Chemical Society, Saratoga Springs, New York **(2003)**, p. 119.
49. M. Torrez, M. Schultenhenrich, and D. R. Livesay, *Biophys. J.* 85, 2845 **(2003)**.
50. S. Kumar, C. J. Tsai, and R. Nussinov, *Protein Eng.* 13, 179 **(2000)**.
51. H. X. Zhou, *Biophys. J.* 83, 3126 **(2002)**.
52. B. N. Dominy, H. Minoux, and C. L. Brooks III, *Proteins* 57, 128 **(2004)**.
53. L. Xiao and B. Honig, *J. Mol. Biol.* 289, 1435 **(1999)**.
54. E. Alsop, M. Silver, and D. R. Livesay, *Protein Eng.* 16, 871 **(2003)**.
55. H. Zhou and F. Dong, *Biophys. J.* 84, 2216 **(2003)**.
56. G. I. Makhatadze, V. V. Loladze, A. V. Gribenko, and M. M. Lopez, *J. Mol. Biol.* 336, 929 **(2004)**.
57. J. Meyer, M. D. Clay, M. K. Johnson, A. Stubna, E. Muenck, C. Higgins, and P. Wittung-Stafshede, *Biochemistry* 41, 3096 **(2002)**.
58. M. Matsumura, W. J. Becktel, M. Levitt, and B. M. Matthews, *Proc. Natl. Acad. Sci. USA* 86, 6562 **(1989)**.
59. A. Cooper, S. J. Eyles, S. E. Radford, and C. M. Dobson, *J. Mol. Biol.* 225, 939 **(1992)**.
60. J. Clarke and A. Fersht, *Biochemistry* 32, 4322 **(1993)**.
61. T. R. Vogl, R. Brengelmann, H. J. Hinz, M. Scharf, M. Lotzbeyer, and J. W. Engels, *J. Mol. Biol.* 254, 481 **(1995)**.
62. R. Guzzi, L. Sportelli, C. La Rosa, D. Milardi, D. Grasso, M. P. Verbeet, and G. W. Canters, *Biophys. J.* 77, 1052 **(1999)**.
63. A. Nemeth, S. Kamondi, A. Szilagyi, C. Magyar, Z. Kovari, and P. Zavodszky, *Biophys. Chem.* 132, 229 **(2002)**.
64. N. Kannan and S. Vishveshwara, *Prot. Eng.* 13, 753 **(2000)**.
65. S. K. Burley and G. A. Petsko, *Science* 229, 23 **(1985)**.
66. A. V. Teplyakov, I. P. Kuranova, E. H. Harutyunyan, B. K. Vainshtein, C. Frommel, W. E. Hohne, and K. S. Wilson, *J. Mol. Biol.* 214, 261 **(1990)**.
67. D. J. Squirrell, C. R. Lowe, P. J. White, and J. A. H. Murray, *PCT Int. Appl.* **(1996)**, p. 40.
68. D. Perl, U. Mueller, U. Heinenmann, and F. X. Schmid, *Nature Struct. Biol.* 7, 380 **(2000)**.
69. M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. G. M. van Loon, and M. Wyss, *Protein Eng.* 15, 403 **(2002)**.
70. C. S. Piddington, C. S. Houston, M. Paloheimo, M. Cantrell, A. Miettinen-Oinonen, H. Nevalainen, and J. Rambosek, *Gene* 133, 55 **(1993)**.
71. D. B. Mitchell, K. Vogel, B. J. Weimann, L. Pasamontes, and A. P. M. G. van Loon, *Microbiology* 143, 245 **(1997)**.
72. A. Crameri, S. A. Raillard, E. Bermudez, and W. P. C. Stemmer, *Nature* 391, 288 **(1998)**.
73. M. S. Dennis, A. Herzka, and R. A. Lazarus, *J. Biol. Chem.* 270, 25411 **(1995)**.
74. B. I. Dahiyat and S. L. Mayo, *Science* 278, 82 **(1997)**.
75. N. Amin, A. D. Liu, S. Ramer, W. Achle, D. Meijer, M. Metin, S. Wong, P. Gualfeti, and V. Schellenberger, *Protein Eng. Design and Selection* 17, 787 **(2004)**.
76. M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes, and A. P. G. M. van Loon, *Protein Eng.* 13, 49 **(2000)**.
77. X. Jiang, E. J. Bishop, and R. S. Farid, *J. Am. Chem. Soc.* 119, 838 **(1997)**.
78. V. Renugopalakrishnan, A. Strzelczyk, P. Li, A. A. Mokhnatyuk, S. H. Gursahani, M. Nagaraju, M. Prabhakaran, H. Arjomandi, and S. L. Lakka, *Int. J. Quantum Chem.* 95, 627 **(2003)**.
79. R. Moreno, O. Zafra, F. Cava, and J. Berenguer, *Plasmid* 49, 2 **(2003)**.
80. J. E. Wampler and E. A. Neuhaus, *J. Prot. Chem.* 16, 727 **(1997)**.
81. R. Hiller, Z. H. Zhou, M. W. W. Adams, and S. W. Englander, *Proc. Natl. Acad. Sci. USA* 94, 11329 **(1997)**.
82. R. F. Blake, J. B. Park, Z. H. Zhou, D. R. Hare, M. W. Adams, and M. F. Summers, *Protein Sci.* 1, 1522 **(1992)**.
83. M. K. Eidsness, K. A. Richie, A. E. Burden. D. M. Kurtz, Jr., and R. A. Scott, *Biochemistry* 36, 10406 **(1997)**.

84. H. Sticht and P. Rosch, *Prog. Biophys. Mol. Biol.* 70, 95 **(1998)**.
85. T, Devanathan, J. M. Akagi, R. T. Hersh, and R. H. Himes, *J. Biol. Chem.* 244, 2846 **(1969)**.
86. J. M. Chandonia, G. Hon, N. S. Walker, L. Conte, P. Koehl, M. Levitt, and S. E. Brenner, *Nucl. Acids Res.* 32, D189 **(2004)**.
87. T. Milledge, S. Khuri, X. Wei, C. Yang, G. Zheng, and G. Narasimhan, in *Proc. of Intl. Symp. on Computational Biology and Genome Informatics* (*CBGI*), Salt Lake City, UT **(2005)**.
88. I. N. Berezovsky and E. I. Sakhhnovich, *Quantitative Biology*, Los Alamos National Laboratory, Preprint Archive **(2004)**, p. 1.
89. F. Tekaia, E. Yeramian, and B. Dujon, *Gene* 297, 51 **(2002)**.
90. B. Dalhus, M. Saarinen, U. H. Sauer, P. Eklund, K. Johansson, A. Karlsson, S. Ramaswamy, A. Bjork, B. Synstad, K. Naterstad, R. Sirevag, and H. Eklund, *J. Mol. Biol.* 318, 707 **(2002)**.
91. S. Kumar, C. J. Tsai, B. Ma, and R. Nussinov, *J. Biomol. Struct. Dyn. Conversation* 11, 79 **(2000)**.
92. A: M. Facchiano, G. Colonna, and R. Ragone, *Prot. Eng.* 11, 753 **(1998)**.
93. C. Vieille and G. J. Zeikus, *Microbiol. Mol. Biol. Rev.* 65, 1 **(2001)**.
94. G. Gianese, P. Argos, and S. Pascarella, *Prot. Eng.* 14, 141 **(2001)**.
95. J. P. Chessa, I. Petrescu, M. Bentahir, J. V. Beeumen, and C. Gerday, *Biochim. Biophys. Acta* 1479, 265 **(2000)**.
96. G. Kleiger, R. Grothe, P. Mallick, and D. Eisenberg, *Biochemistry* 41, 5990 **(2002)**.
97. D. La, M. Silver, R. C. Edgar, and D. R. Livesay, *Biochemistry* 42, 8988 **(2003)**.
98. P. A. Fields, *Comp. Biochem. Physiol. A* 129, 417 **(2001)**.
99. G. van Driessche, W. Hu, G. van de Werken, F. Selvaraj, J. D. McManus, R. E. Blankenship, and J. J. van Beeumen, *Prot. Sci.* 8, 947 **(1999)**.
100. N. Haulinen, O. Turunen, J. Janis, M. Leisola, and J. Rouvinen, *Eur. J. Biochem.* 270, 1399 **(2003)**.

Received: 28 July 2005. Accepted: 29 July 2005.