JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

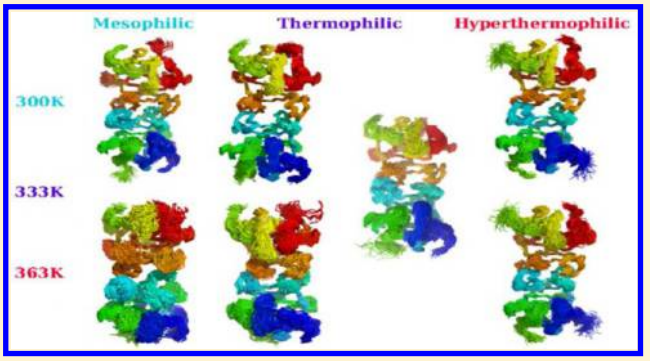
Molecular Dynamics Perspective on the Protein Thermal Stability: A Case Study Using SAICAR Synthetase

Kavyashree Manjunath and Kanagaraj Sekar*

Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, Karnataka 560 012, India

Supporting Information

ABSTRACT: The enzyme SAICAR synthetase ligates aspartate with CAIR (5′-phosphoribosyl-4-carboxy-5-aminoimidazole) forming SAICAR (5-amino-4-imidazole-N-succinocarboxamide ribonucleotide) in the presence of ATP. In continuation with our previous study on the thermostability of this enzyme in hyper-/thermophiles based on the structural aspects, here, we present the dynamic aspects that differentiate the mesophilic (*E. coli*, *E. chaffeensis*), thermophilic (*G. kaustophilus*), and hyperthermophilic (*M. jannaschii*, *P. horikoshii*) SAICAR synthetases by carrying out a total of 11 simulations. The five functional dimers from the above organisms were simulated using molecular dynamics for a period of 50 ns each at 300 K, 363 K, and an additional simulation at 333 K for the thermophilic protein. The basic features like root-mean-square deviations, root-mean-square fluctuations, surface accessibility, and radius of gyration revealed the instability of mesophiles at 363 K. Mean square displacements establish the reduced flexibility of hyper-/thermophiles at all temperatures. At the simulations time scale considered here, the long-distance networks are considerably affected in mesophilic structures at 363 K. In mesophiles, a comparatively higher number of short-lived (having less percent existence time) Cα, hydrogen bonds, hydrophobic interactions are formed, and long-lived (with higher percentage existence time) contacts are lost. The number of time-averaged salt-bridges is at least 2-fold higher in hyperthermophiles at 363 K. The change in surface accessibility of salt-bridges at 363 K from 300 K is nearly doubled in mesophilic protein compared to proteins from other temperature classes.



INTRODUCTION

Life on the earth has evolved to survive various environmental conditions like temperature, pressure, salt concentrations, pH, etc. Based on the temperature of survival, organisms have been classified as psychrophiles, mesophiles, thermophiles, and hyperthermophiles. Psychrophiles thrive at −15 to 10 °C, mesophiles thrive at normal temperatures of 20 to 45 °C, thermophiles can survive at 45 °C up to 80 °C, while hyperthermophiles inhabit an extreme temperature greater than or equal to 80 °C and sometimes endure more than 100 °C. The organisms surviving at extremely high temperature conditions,[1] where mesophilic organisms generally perish, have come up with various strategies of adaptations[2] at cellular and molecular levels. At the cellular level, the cell membranes are kept intact due to the presence of ether linkages, longer hydrocarbon chains [i.e., higher caldarchaeol (C_{40}) to archaeol (C_{20}) ratio], and increased degree of cyclization in hyperthermophiles.[2] At the molecular level, DNA and RNA are protected from degradation at high temperatures due to the presence of high concentrations of linear or branched chain polyamines. In addition, DNA is protected by a positive superhelix provided by the reverse gyrase enzyme that is found only in hyperthermophiles.[2] Proteins in hyperthermophiles are extremely stable although they are composed of the same set of 20 naturally occurring amino acids and have considerable sequence and structural identities with their mesophilic counterparts. These proteins have different strategies for adapting to higher temperature. The study on thermostability of proteins has been of interest since Perutz work on thermal stability of ferridoxins[3] because of its relevance to industrial applications[4] and protein folding.

Many aspects relating to the sequence, structure, and dynamics contributing to the protein stability in thermophilic and hyperthermophilic organisms have been explored. At very high temperatures, the noncovalent interactions like hydrophobic, hydrogen bonds, and ionic bonds holding the structure together breaks down leading to unfolding of the protein. Subsequently, the unfolded protein denatures due to the deamidation of Asn and Gln residues,[5,6] hydrolysis of peptide bonds, β-elimination of cystine residues, and disulfide interchange.[6] Thus, the proteins present in hyperthermophiles protect their structure by increasing the noncovalent interactions and avoiding the exposure of susceptible residues to solvent. Studies done on protein sequences indicate preferential amino acid substitutions in hyperthermophilic

Received: May 22, 2013
Published: August 20, 2013

organisms with a greater tendency toward higher hydrophobic and charged residues and lesser polar uncharged residues.[7,8] Initial statistical studies indicated a preferential substitution of Gly→Ala and Lys→Arg,[9] and a decrease in β-branched residues like Ile, Val, and Thr.[10] The composition of Arg residue increases from mesophiles to thermophiles due to the more charged interactions it provides, reduced activity of guanidinium moiety, and its capacity to maintain net positive charge at elevated temperatures.[4] Some proteins contain cysteines that form disulfide bonds which stabilize the protein by reducing the entropy of the unfolded state.[11] However, all the above stated reasons are not observed universally in all hyper-/thermophilic proteins.[4]

Structural factors contributing to the thermostability of these proteins include shorter surface loops,[12] anchoring of loops,[4] presence of disulfide bonds,[13] increased hydrophobic[14−17] and aromatic interactions,[18,19] higher number of ion pairs,[4,20−24] enhanced hydrogen bonding,[4,7,25−27] cation-π interactions,[28,29] increased packing density,[30] lesser number of cavities, higher rigidity,[31] and higher oligomerization states.[32] All of the above-mentioned properties need not necessarily be present in a given thermostable protein; however, combinations of one or more of these properties contribute to the thermostability.[4] Several studies on thermodynamical aspects of thermal stability have also been carried out on proteins from mesophiles and hyper-/thermophiles.[33−37]

Molecular dynamic simulations done on mesophilic and hyper-/thermophilic proteins revealed a few distinguishing features. At the secondary structure level, β-sheets are found to be more stable than the α-helices, as shown by the simulations of β-lactamases.[38] High temperature denaturation simulations of family 11 xylanases revealed that N-terminal and α-helices initiate the unfolding process.[39] The thermophiles exhibit an extensive network of water−water hydrogen bonds in the hydration shell of the protein thereby protecting the protein scaffold.[40] A relatively higher number of exposed hydrophobic patches in mesophiles on the protein surface destabilizes the water−water hydrogen bonding network.[41] Another MD study indicates the presence of a higher number of salt bridges and their clustering in hyperthermophilic proteins.[42] One of the important aspects regarding the mesophilic and hyper-/thermophilic proteins is their flexibility. The concept of corresponding states suggests that these proteins show a similar degree of flexibility in their respective growth temperatures.[43] In contrary to the above, a different scenario is observed in another study.[40] However, a recent study involving 19 pairs of homologous mesophilic and thermophilic enzymes using constraint network analysis concluded that the thermophilic enzymes exhibit flexibility at the active site, while maintaining rigidity in other regions in order to maintain thermal stability.[44]

In the present work, the enzyme phosphoribosylaminoimidazole-succinocarboxamide (SAICAR) synthetase from mesophilic, thermophilic, and hyperthermophilic organisms are considered for studying the thermostability based on the dynamics of proteins at different temperatures. The enzyme catalyzes the seventh step in the purine biosynthesis pathway by ligating 5′-phosphoribosyl-4-carboxy-5-aminoimidazole (CAIR) with aspartate in an ATP dependent reaction. It is a monofunctional enzyme in archaea, bacteria, fungi, and plants but bifunctional in higher eukaryotes incorporating another domain with AIR carboxylase activity. The oligomeric state of the monofunctional SAICAR synthetase is reported to be a monomer in *S. cerevisiae*,[45] a noncovalent dimer in *E. coli*[46] and *P. horikoshii*,[47] and a covalent dimer in *T. maritime*.[48] According to SCOP classification, the enzyme belongs to the α+β class, with the β-sheets forming the platform for binding of the substrates and the longest α-helix forms a backbone for the active site platform, while other helices (α and 3_{10}) form the ridges supporting the active site. The sequence of *P. horikoshii* SAICAR synthetase has 41.7%, 42.7%, 49.4%, and 49.6% identity with *E. coli, E. chaffeensis, G. kaustophilus*, and *M. jannaschii* sequences, respectively, and all have more than 60% sequence similarity with that of *P. horikoshii*. The three-dimensional structure of *P. horikoshii* SAICAR synthetase has an rmsd (root-mean-square deviation) of less than 1.36 Å with the structures from above-mentioned organisms. Despite significant sequence and structural similarities, there were few distinctions between the mesophilic, thermophilic, and hyperthermophilic enzymes,[47] but the structures have limitations as they are one among many conformations it can take at a given condition and the configuration of the structure depends to some extent on the crystallization condition used and temperature at which the data was collected. Thus, it would be more definitive to study the dynamic behavior of protein under identical conditions for thermostability. The present work is an effort to extract more deterministic features of thermal stability (hyper-/thermophiles) or susceptibility (mesophiles) from its dynamic nature much before it starts denaturing or during the initial stages of denaturation.

MATERIALS AND METHODS

Protein Models. The atomic coordinates of the protein structures were downloaded from PDB. The structure of SAICAR synthetase from *E. coli* (PDB-id: 2gqr), *E. chaffeensis* (PDB-id: 3kre), *G. kaustophilus* (PDB-id: 2ywv), *M. jannaschii* (PDB-id: 2z02), and *P. horikoshii* (PDB-id: 3u55) were considered for simulation, and these structures are hereafter referred as *Ec*SS, *Eh*SS, *Gk*SS, *Mj*SS, and *Ph*SS, respectively, with the simulation temperature indicated in the subscript. Among the above structures, *Ec*SS and *Eh*SS were from mesophilic, *Gk*SS was from thermophilic, and *Mj*SS and *Ph*SS were from hyperthermophilic organisms. Missing side chains and the residues at the N-terminus of *Eh*SS, *Gk*SS, and *Ph*SS were modeled using the server SWISS-MODEL[49] and COOT[50] using the closest homologue for which the full length structure was available. Functional dimers (similar to the asymmetric unit of *Ec*SS) were considered for all simulations. For those structures with a monomer in the asymmetric unit, functional dimers were generated using symmetry equivalent molecules. All heteroatoms including water oxygen atoms were removed before starting the simulation.

Molecular Dynamics. Simulations were carried out using the GROMACS v4.5.3[51] package with OPLS-AA (optimized potentials for liquid simulations all atom) force field[52] and TIP4P water model.[53] Proteins were simulated in a water filled dodecahedron box with a distance of 1.2 nm between the protein and the box. Standard protonation states of the residues were used. The charges of the system were neutralized by the addition of either Na^+ or Cl^- ions. The electrostatic interactions were treated using the Particle Mesh Ewald (PME)[54] method, with a coulomb cutoff of 1.4 nm, a Fourier spacing of 0.12 nm with a fourth order interpolation. The van der Waals interactions were treated using switch potential with a cutoff of 1.0 nm with the switch function applied from 0.9 nm. The bond lengths were constrained using the LINCS

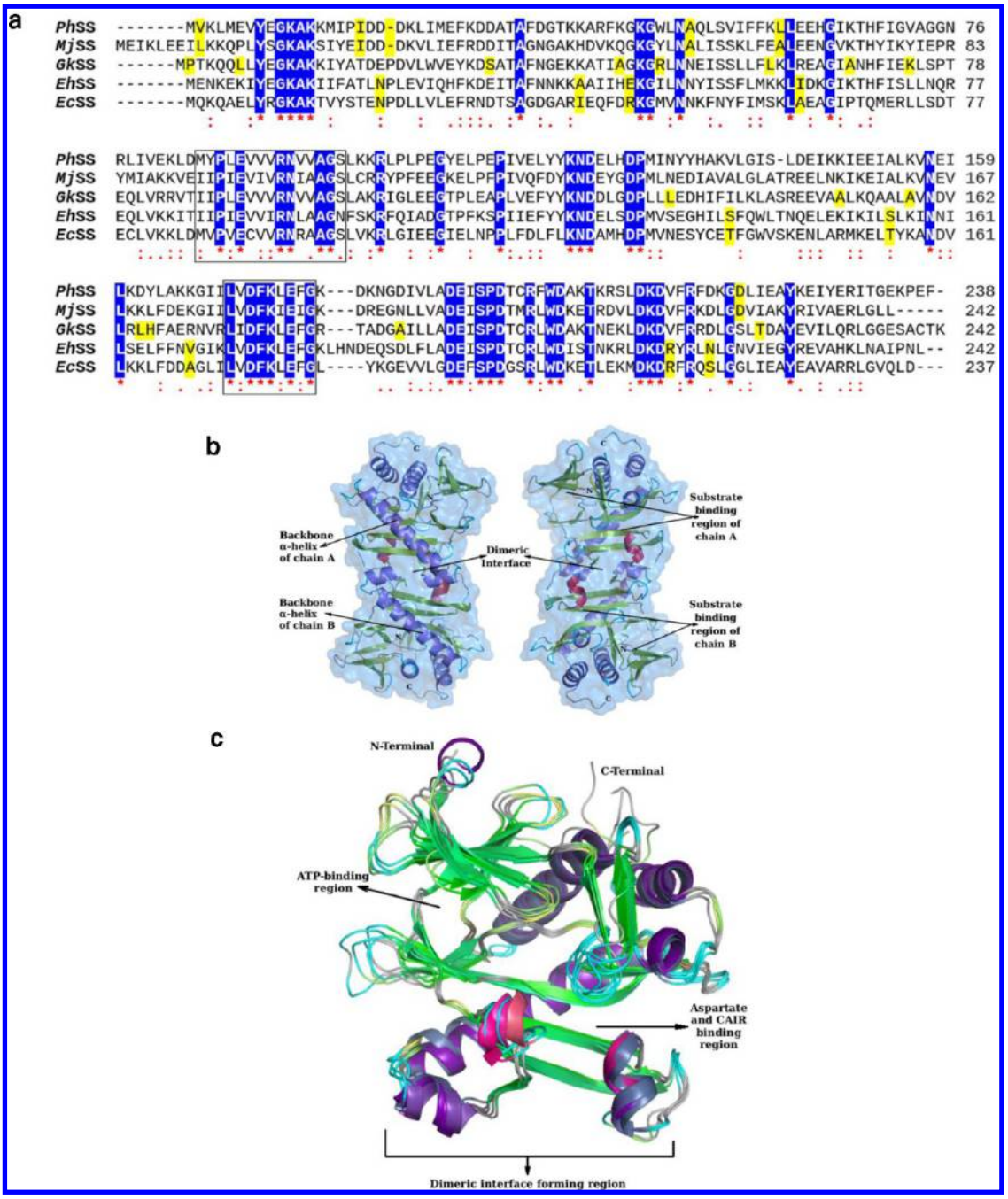


Figure 1. (a) Multiple sequence alignment of SAICAR synthetase sequences from *Ph*SS, *Mj*SS, *Gk*SS, *Eh*SS, and *Ec*SS. Conserved residues are highlighted in blue, residues which are uniquely different in mesophilic, thermophilic, and hyperthermophilic proteins are highlighted in yellow. Residues mentioned inside box are the SAICAR synthetase signature sequences. (b) The two views of dimeric quaternary orientation of SAICAR synthetase using which all the simulations were carried out. Some important regions are labeled. (c) Structural alignment of monomeric unit of all five SAICAR synthetase, secondary structures are colored with similar colors.

algorithm,[55] and the energy minimization was carried out to remove all short contacts using conjugate gradient algorithm with a convergence criteria of 1 kJ mol^{-1} nm^{-1}. Following energy minimization, temperature and pressure equilibration were performed by applying the position restraints on the protein. The system was coupled to a temperature bath using the Nosé-Hoover[56,57] thermostat with a coupling time constant, τ_t, of 0.4 ps. Initial velocities at an appropriate temperature (300 K/333 K/363 K) were generated from the Maxwell–Boltzmann distribution using a random number generator. The pressure of the system was isotropically (compressibility of 4.5 × 10^{-5} bar^{-1}) coupled to a barostat at 1 bar using the Parrinello-Rahman method[58] with a coupling constant, τ_p, of 2 ps. The position restrained dynamics was run for 100 ps during both temperature and pressure equilibration by applying a force constant of 1000 kJ mol^{-1} nm^{-2} on protein atoms. The simulations were run with a time step of 2 fs with the atomic coordinates and velocities saved for every 2 ps. All production simulations were run for 50 ns. The structures *Ec*SS, *Eh*SS, *Gk*SS, *Mj*SS, and *Ph*SS were simulated at 300 K and 363 K. In addition, *Gk*SS was also simulated at 333 K. The total number atoms used in the simulations were 199704, 153192, 156406,

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448–2461

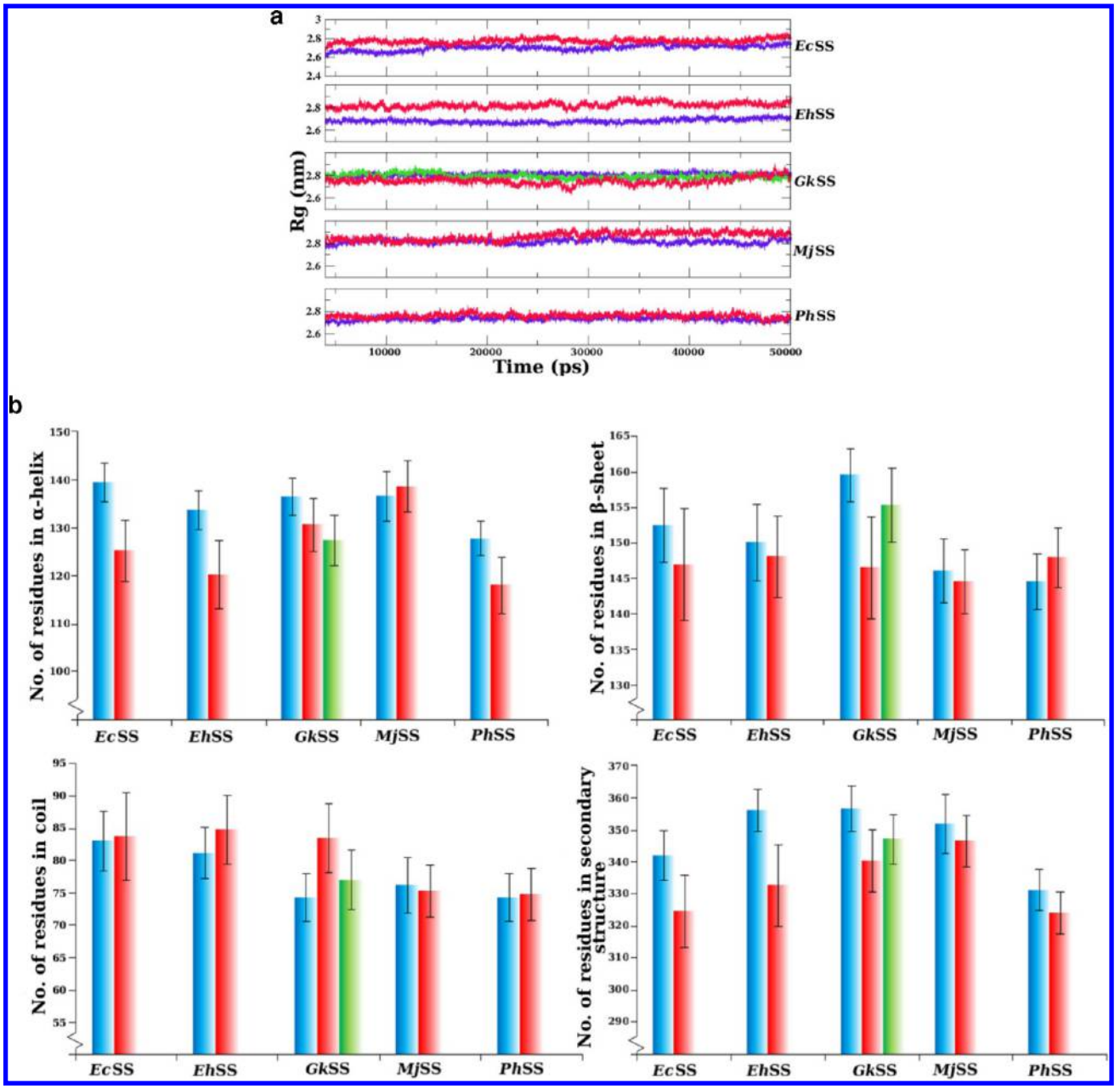


Figure 2. (a) Plot of radius of gyration (R_g) of protein along the trajectory. Plots at 300 K, 333 K, and 363 K are colored in blue, green, and red, respectively. (b) Time averaged and standard deviations of number of residues in different secondary structures of the protein over the entire trajectory. Blue, green, and red bars correspond to the respective secondary structures at 300 K, 333 K, and 363 K, respectively.

198138, and 206950 for *Ec*SS, *Eh*SS, *Gk*SS, *Mj*SS, and *Ph*SS, respectively. All the simulations are analyzed starting from 4 ns to 50 ns.

Alignment of the sequences was done using ClustalW,[59] and structural superposition was carried out using mustang.[60] Before the start of the analysis, it was made sure that the dimer of the protein was intact throughout the simulations, and jumps of the molecules out of the box were removed. The simulation trajectories were analyzed using the tools available in the Gromacs package. The program DSSP[61] was used to assign the secondary structures of the proteins. The evolution of the secondary structural content along the trajectory was calculated by time-averaging the number of residues in each secondary structure of the protein with *do_dssp* as an interface. The rmsf (root-mean-square fluctuation) plots were integrated by calculating the area under the curve and normalizing the values by dividing them by the lengths of the corresponding protein sequence and multiplying by 100. This value is termed as PnA_{rmsf} (Percentage of normalized Area of rmsf curve) in the text. The trajectory of the protein structures superposed on the first structure was generated for the mean square displacement (msd) calculations. The msd was calculated for Cα atoms based on the following formula

$$\mathrm{msd}(\Delta t) = \frac{1}{N(T-\Delta t+1)} \sum_{i}^{N} \sum_{t=0}^{T-\Delta t} [R_i(t+\Delta t) - R_i(t)]^2$$

where N is the total number of atoms, T is the total time of trajectory (here it is 46 ns, as the first 4 ns have been

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448−2461

Table 1. Time-Averages and Standard Deviation of rmsd, PnA_{RMSF}, and sasa

	rmsd[a]		PnA_{RMSF}[b]		solvent accessible surface area (sasa)		
	A	B	A	B	charged	hydrophilic	hydrophobic
$EcSS_{300}$	0.17 ± 0.02	0.21 ± 0.06	7.47	9.90	125.73 ± 2.11	56.44 ± 0.88	71.87 ± 2.39
$EcSS_{363}$	0.34 ± 0.08	0.27 ± 0.04	14.11	11.97	123.15 ± 2.00	55.07 ± 1.56	80.98 ± 2.45
$EhSS_{300}$	0.17 ± 0.04	0.12 ± 0.01	9.32	6.59	96.42 ± 2.37	80.14 ± 1.20	84.49 ± 1.85
$EhSS_{363}$	0.27 ± 0.04	0.23 ± 0.07	12.86	13.34	103.92 ± 2.81	81.29 ± 1.68	93.67 ± 2.48
$GkSS_{300}$	0.23 ± 0.02	0.20 ± 0.03	8.46	7.24	122.11 ± 1.66	56.74 ± 0.94	77.52 ± 1.45
$GkSS_{333}$	0.24 ± 0.04	0.23 ± 0.03	9.61	8.73	119.77 ± 1.99	58.73 ± 1.17	79.19 ± 2.59
$GkSS_{363}$	0.21 ± 0.03	0.26 ± 0.05	10.77	11.28	116.46 ± 2.10	60.80 ± 1.29	81.40 ± 1.99
$MjSS_{300}$	0.14 ± 0.02	0.17 ± 0.03	7.72	6.59	146.28 ± 1.85	46.65 ± 0.98	67.20 ± 1.62
$MjSS_{363}$	0.20 ± 0.03	0.17 ± 0.02	9.92	8.98	143.86 ± 2.09	47.12 ± 1.03	73.42 ± 2.06
$PhSS_{300}$	0.19 ± 0.03	0.21 ± 0.03	8.50	7.60	138.43 ± 1.95	42.39 ± 0.81	78.70 ± 1.68
$PhSS_{363}$	0.22 ± 0.02	0.25 ± 0.03	9.72	10.89	135.36 ± 2.17	42.98 ± 1.01	84.45 ± 2.16

[a]Average ± standard deviation of rmsd. [b]Percentage of normalized area of rmsf curve [(Integral of the curve)/(length of protein)*100].

discarded), Δt is the time between restarts (here 2 ps), and R_i is the coordinates of the atoms considered. The msd is calculated for definite intervals of time (2, 4, 6, 10 ps, etc.) during the whole trajectory considering regularly spaced time origins. Thus, although all the calculations are started from 4 ns onward to 50 ns, the actual msd plot will be from 0 ns to 46 ns.

The Cα contacts or hydrophobic contacts were calculated without double counting a contact. Hydrogen bonds were calculated using a maximum cutoff distance between the donor and acceptor as 3.5 Å and the donor-hydrogen-acceptor angle to be ≥120°. The hydrogen bonds formed between the same acceptor and donor pair but different hydrogen atoms were merged as a single hydrogen bond. The salt-bridges (SBs) were calculated between the atoms of negatively charged residues Asp (Oδ1, Oδ2), Glu (Oε1, Oε2), and positively charged residues Arg (Nε, Nη1, Nη2), Lys (Nζ), with a cutoff distance of 4 Å (short-range) to 7 Å (long-range). Histidine and terminal carboxylic and amide groups were ignored while calculating the SBs. For some calculations, the PERL and AWK scripts were used. The figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). The plots were generated using Xmgrace (Paul J. Turner Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology Beaverton, Oregon) and Microsoft Office PowerPoint 2007.

RESULTS AND DISCUSSION

According to the SCOP assignments, SAICAR synthetase belongs to the $\alpha+\beta$ class of proteins. The secondary structures of five proteins considered in this study have 26%−30% α-helices and 30%−33% β-strands. The longest α-helix form the backbone of the enzyme and other helices either guard the entrance of the active site or form the dimeric interface. The β-strands mainly form the binding regions for the substrates ATP, CAIR, and aspartate. Multiple sequence alignment of these five proteins shows significant similarity as illustrated in Figure 1a. The SAICAR synthetase signature sequences are shown enclosed in the box. The regions identical in all sequences are highlighted in blue, while the regions which differ uniquely in hyper-/thermophiles or mesophiles are highlighted in yellow. *Ph*SS and *Mj*SS have four out of five positions substituted with nonpolar residues in place of polar residues found in other sequences. The *Gk*SS has nine out of 15 positions substituted with a nonpolar residue, where polar residue occurs in other sequences. Our previous study[47] also showed that *Gk*SS has increased nonpolar residues. Five out of nine (polar/nonpolar) positions in mesophilic (*Ec*SS and *Eh*SS) sequences have been substituted with charged residues in other sequences. In summary, the hyper-/thermophiles have more nonpolar and charged residues compared to the mesophilic counterparts. The functional unit of the enzyme is a dimer (Figure1b). The superposition of the monomeric units shows a phenomenal similarity among the structures, as shown in Figure 1c, with all the active site residues being structurally conserved. There is no shortening of the loop in these hyper-/thermophilic proteins as opposed to the observations made in other studies.[12] In the present work, we extend the structural analysis[47] of thermostability of SAICAR synthetase using molecular dynamics at different temperatures. A total of 11 simulations (50 ns each) are performed on the functional dimers of the above-mentioned SAICAR synthetase structures from mesophilic (*Ec*SS and *Eh*SS), thermophilic (*Gk*SS), and hyperthermophilic (*Mj*SS and *Ph*SS) organisms at 300 K and 363 K including the simulation of *Gk*SS at 333 K. The overall structural topology is more or less maintained in all cases during the time scale of simulation, but subtle characteristic changes are observed delineating mesophilic, thermophilic, and hyperthermophilic proteins.

Preliminary Analysis of Trajectory. The compactness of the protein along the trajectory is analyzed by calculating the radius of gyration (R_g) of the protein using g_gyrate weighted by atomic mass (Figure 2a). It can be observed that the R_g reasonably differ between $EcSS_{300}$/$EhSS_{300}$ and $EcSS_{363}$/$EhSS_{363}$, while thermophilic (*Gk*SS) and hyperthermophilic proteins (*Mj*SS and *Ph*SS) hardly show any changes. Thus, the increase in the R_g indicates expanding protein structure, probably due to the loosening of the structural network at 363 K. A gentle variation in the secondary structures (calculated using DSSP) is observed along the trajectory between 300 K and 363 K for all proteins. Overall secondary structural content, (α-helix+β-sheet+β-bridge+turn), decreased more predominantly in mesophiles from 300 K to 363 K. The secondary structural contents of the two chains are differentially affected in all the simulations. Figure 2b shows the average and standard deviations of the evolution of the overall secondary structure, α-helix, β-sheets, and coil along the trajectory. In mesophiles, α-helices appear more sensitive to temperature compared to β-sheets. *Mj*SS appears to increase its helical content with temperature, but, considering the standard deviation, it is not significant. The breakage of hydrogen bonds is probably the main cause for the decrease in the secondary structural content. Thus, an overall change in the

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448−2461

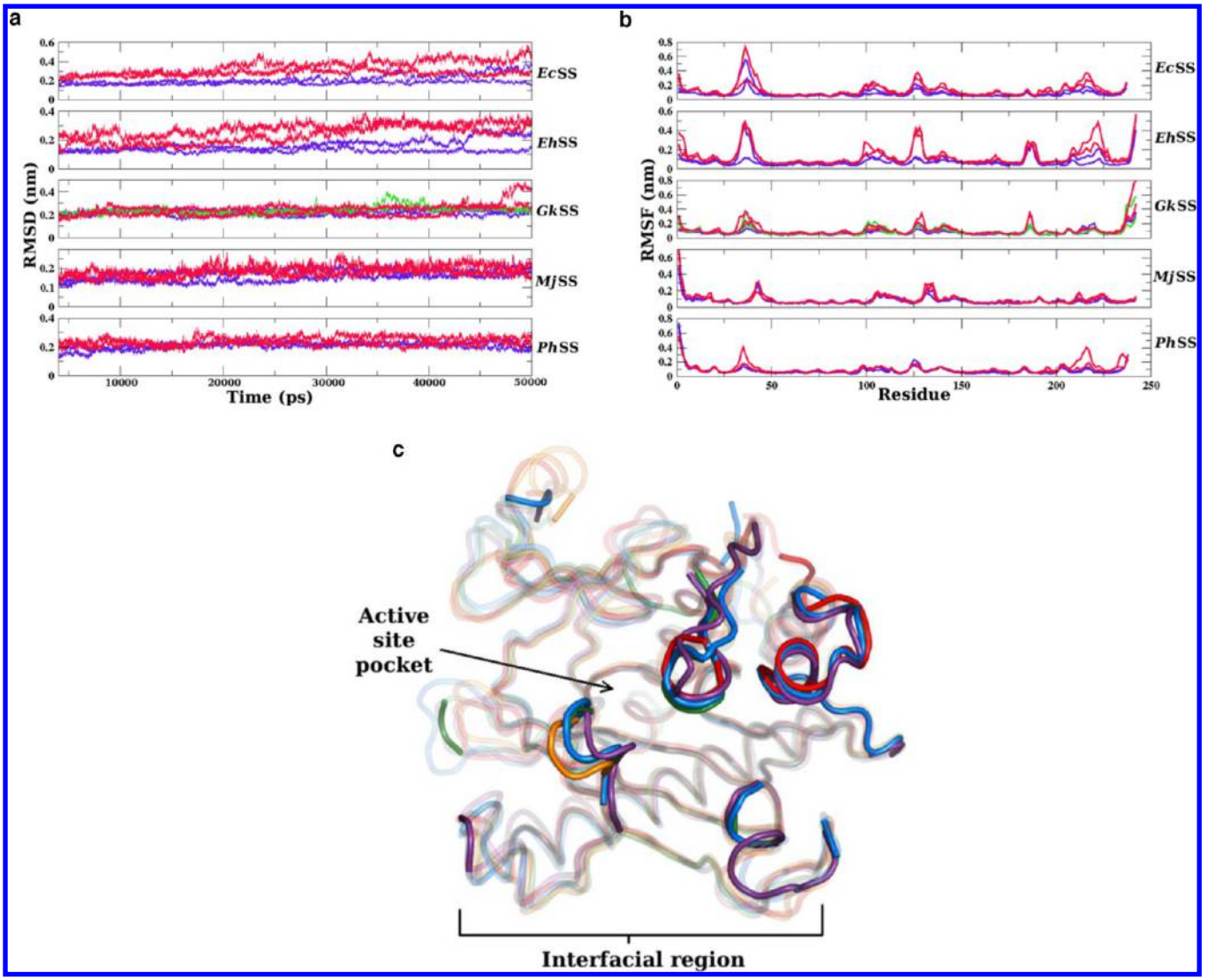


Figure 3. (a) The Cα rmsd plots of each chain of all the proteins at 300 K (blue), 333 K (green), and 363 K (red). (b) The backbone rmsf plots of two chains in all proteins over the trajectory. Graphs are colored similar to rmsd plot. (c) Regions corresponding to a difference of the RMSF value more than 0.1 nm between 363 K/333 K and 300 K trajectories. Structures in blue shades correspond to mesophilic proteins and orange/red corresponds to hyperthermophilic proteins while those in green correspond to thermophilic protein.

secondary structural content differentiates mesophiles from hyper-/thermophiles.

The unfolding of proteins would expose the buried hydrophobic residues to the water environment. This is examined by measuring the solvent accessible surface area (sasa) of the whole protein and for each group of charged (Asp, Glu, Lys, and Arg), hydrophilic (Ser, Thr, Asn, Gln, Gly, His, and Pro), and hydrophobic residues (Ala, Val, Leu, Ile, Met, Cys, Phe, Tyr, and Trp).[62] The Supporting Information Table S1 lists some results from sasa calculations. The first column shows the difference between sasa of proteins at 363 K/333 K and corresponding sasa at 300 K . The second column specifies the sum of the area of only those residues which get exposed at 363 K/333 K. An appreciable difference can be seen in these values between mesophilic/thermophilic (*Ec*SS, *Eh*SS, and *Gk*SS at 363 K) and hyper-/thermophilic proteins (*Gk*SS at 333 K and *Mj*SS, *Ph*SS at 363 K). These values indicate that *Gk*SS and the mesophilic proteins at 363 K have similar sasa, while *Gk*SS at 333 K and hyperthermophilic proteins at 363 K have comparable accessibilities. Due to the lesser noncovalent interactions in mesophiles compared to hyperthermophiles (see later sections), the susceptible residues get exposed more easily at 363 K, while hyperthermophiles on the other hand have a higher number of nonbonded interactions. Further, the hyperthermophiles (at 363 K) and thermophiles (at 333 K) have a larger number of atoms completely buried (with zero sasa) compared to mesophilic proteins (Supporting Information Table S1), providing them with additional capacity to withstand higher temperatures through hydrophobic interactions. Table 1 shows the time averaged values of the sasa with standard deviations. In correlation with the amino acid composition, the hyperthermophilic proteins contain a marginally lesser accessible surface area of hydrophilic residues and a higher accessible surface area of charged residues compared to mesophilic proteins. The hydrophilic amino acids more or less retain a similar accessible surface area at both the temperatures in all the proteins. The hydrophobic accessible surface area increases relatively more in mesophiles compared to hyperthermophiles at 363 K. Accessibility of charged residues, except that of *Eh*SS, reduces at higher temperatures. Thus, at the time

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448−2461

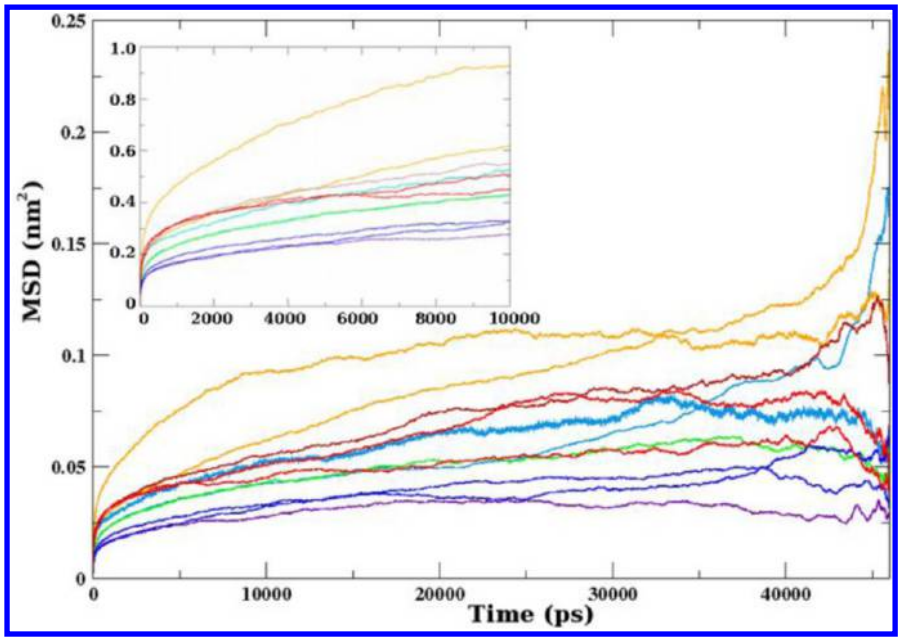


Figure 4. Mean square displacement (msd) plots. The msds of mesophiles, thermophile, and hyperthermophiles at 300 K and 363 K are colored light blue, purple, dark blue and orange, dark red and light red respectively, while msds of thermophile at 333 K is colored green. The inset graph shows msds till 10 ns.

scales of simulation considered in this study, only certain aspects of surface accessibility establishes the susceptibility of the mesophilic protein at higher temperature.

RMSD and Fluctuation Analysis. The rmsd (root mean squared deviation) of the Cα atoms in the trajectory, calculated with respect to the initial structure, stabilizes approximately near 4 ns in all proteins. Inspection of rmsd plots (Figure 3a), of both the chains in each protein, indicates that the two chains behave slightly differently. Nonetheless, both the chains show significant escalation in the rmsd at 363 K in the case of mesophilic proteins, which is clearly visible even at the initial stages of simulation. The mean and the standard deviation of rmsd are given in Table 1. The rmsd of *Gk*SS$_{333}$, on the other hand, almost overlaps with that of *Gk*SS$_{300}$, but in the case of *Gk*SS$_{363}$, the rmsd of one chain deviates significantly compared to that of *Gk*SS$_{300}$. The hyperthermophilic proteins have negligible rmsd variation between 300 K and 363 K.

The fluctuations of the residues are analyzed by calculating the root mean squared fluctuation (rmsf) of the backbone atoms, illustrated in Figure 3b. As expected, the mesophilic structures fluctuate higher at 363 K than at 300 K (in certain regions) compared to thermophilic or hyperthermophilic proteins. The PnA$_{rmsf}$ values given in Table 1 demarcates the mesophilic proteins (from hyper-/thermophilic proteins) which almost doubles its value at 363 K. It could be observed from Figure 3b that only some parts of the proteins show a significant increase in the rmsf values, and the fluctuations are asymmetrical in two chains. The regions having a difference of rmsf more than 0.1 nm between 300 K and 363 K trajectories (of individual chains) are filtered. As shown in Figure 3c, these regions cover the active site of the enzyme and occur at the dimeric interface. Even at 300 K , these residues have considerably higher rmsf values than other residues; indicating that most probably these movements are required during substrate binding and product release. Thus, the effect of temperature is first manifested in these regions more significantly in mesophiles and perhaps is the nucleation sites for unfolding.

It is thus clear from the rmsf studies that the thermophilic and hyperthermophilic proteins show a lower level of fluctuations at high temperatures compared to the mesophilic proteins. The rmsf measures the average fluctuation of the atoms over the whole trajectory, but in order to see how the protein coordinates change with respect to the initial coordinates at different time intervals, mean square displacements (msd) are found to be useful. The msd gives the behavior of the molecule (with respect to the structure at the nearest time origin) during different intervals of time. The msd for a given time interval is the average of the squared deviation of the structures from the structure at the nearest origin. Figure 4 shows the msd calculated for all the trajectories. The msd of mesophilic, thermophilic, and hyperthermophilic proteins at 300 K are colored light blue, purple, and dark blue; at 363 K they are colored orange, dark red, and light red, while the thermophilic protein at 333 K is colored green. The statistics are better only till 10 ns. Thermophilic and hyperthermophilic msd's are smaller than the mesophilic msd at 300 K at all the time intervals. Thus, in the present set, mesophilic proteins are more flexible at 300 K than hyper-/thermophiles at the same temperature. The msd of hyperthermophilic protein at 363 K and thermophilic protein at 333 K matches with the msd of mesophilic protein at 300 K, re-establishing Somero's corresponding states hypothesis.[43] A difference in the flexibility among mesophilic/hyperthermophilic proteins is clear from the plot, implying that although the proteins are similar in structure and function and belong to the same temperature class, there can be moderate to high variation in the flexibility. Mesophilic proteins shows higher sensitivity to the increasing temperature compared to the protein from other two temperature classes. Our observation does not strictly follow the conclusions made in a recent study[63] which states that mesophiles have lower flexibility at higher temperature than thermophilic proteins at time scales below 5 ns. Although in the present study, only one

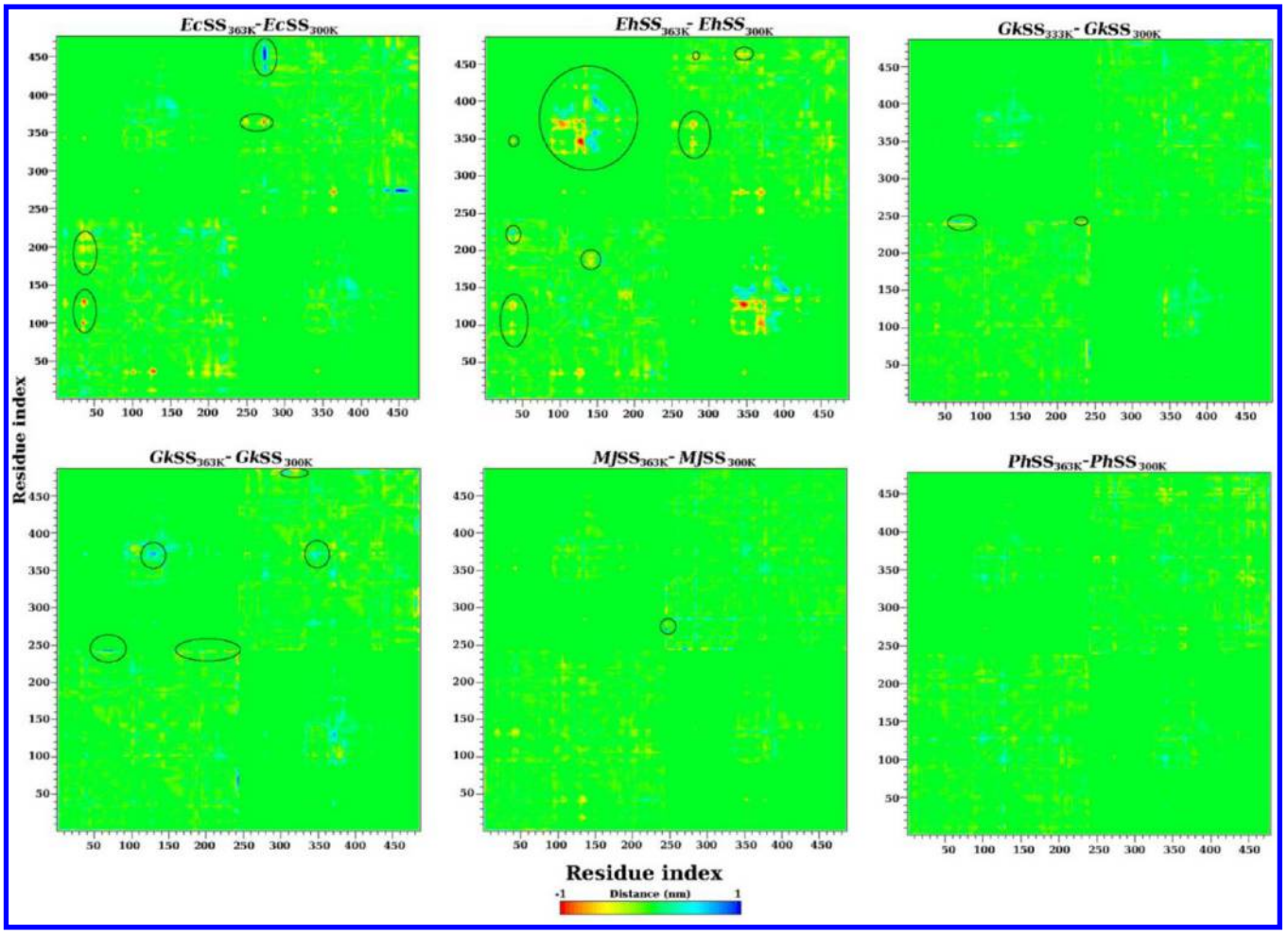


Figure 5. A difference Cα contact maps (based on distance cutoff of 2.5 nm) of proteins over the trajectory. All plots are scaled between −1 nm and +1 nm and colored accordingly as indicated in the scale bar.

of the mesophilic proteins shows lower msd at 363 K. At moderate times scales of simulation, as in the present work, msd calculations serve as an excellent tool in establishing the difference in the dynamics between proteins from different temperature classes.

Contact Analysis. The Cα atom contact map is calculated for the trajectories (4 ns−50 ns) of all proteins with a distance cutoff of 0.4 nm−0.8 nm, 1.0 nm, 1.5 nm, 2.0 nm, and 2.5 nm. The contact maps of 1.0 nm onward are mainly useful for analyzing the changes in the long-range Cα networks in proteins between two temperatures. The difference between the distance matrices at 363 K/333 K and the corresponding matrix at 300 K for each proteins simulation, scaled between −1 nm and 1 nm (difference-distance matrix), are calculated. A distinguishable feature (blue: positive and red: negative regions) in the difference-distance matrix between mesophilic, thermophilic, and hyperthermophilic proteins is observed (Figure 5) clearly in the matrix with a cutoff of 1.5 nm or above. For clarity, the difference-distance matrix with 2.5 nm cutoff is shown in Figure 5. Each map displays the matrix of both the chain with continuous chain numbering, and it is symmetric about the diagonal. The lower left is one chain and the upper right is the other chain, while the upper left and the lower right correspond to the region at the dimeric interface. In *Ec*SS, the differences are visible within each chain, while in *Eh*SS it is more prominent at the dimeric interface. The difference-distance matrix of GkSS$_{363}$-GkSS$_{300}$ has more positive and negative regions compared to GkSS$_{333}$-GkSS$_{300}$. The difference-distance matrices of hyperthermophiles are nearly zero, although, the *Mj*SS map appears to have a negligibly small positive region in one of the chains. Results from the analysis of the short distance (4 Å−8 Å) cutoff Cα contacts are given in Table 2. It lists the gain in the total number of unique contacts of Cα atoms at 363 K/333 K compared to 300 K simulations

Table 2. Cα and Hydrophobic Contact Differences

distance cutoff (Å)	EcSS$_{363}$-EcSS$_{300}$	EhSS$_{363}$-EhSS$_{300}$	GkSS$_{333}$-GkSS$_{300}$	GkSS$_{363}$-GkSS$_{300}$	MjSS$_{363}$-MjSS$_{300}$	PhSS$_{363}$-PhSS$_{300}$
	∑ (Unique Cα Contact Differences)[a]					
4	68	43	25	60	19	32
5	297	256	136	249	138	174
6	368	327	161	318	154	205
7	480	478	237	449	208	321
8	775	681	341	690	337	486
	∑ (Unique Hydrophobic Contact Differences)[b]					
4	9035	10756	4268	7733	5836	8082
5	13687	15454	6541	11802	8628	11769

[a]Sum of all the unique Cα contact differences between 363 K (or at 333 K in case of GkSS) and 300 K. [b]Sum of all the unique hydrophobic contacts differences between 363 K (or at 333 K in case of GkSS) and 300 K.

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448−2461

Table 3. Difference in the Time Averaged Values of HBs between 363/333 K and 300 K

	intraprotein HB[a]	MM HB[b]	SS HB[c]	MS HB[d]	chr-chr HB[e]	phi-phi HB[f]	pho-pho HB[g]
$EcSS_{363}$-$EcSS_{300}$	−7.97	−13.30	1.18	4.49	6.14	−1.84	−1.97
$EhSS_{363}$-$EhSS_{300}$	−18.07	−18.10	0.60	0.00	1.38	2.25	−7.21
$GkSS_{333}$-$GkSS_{300}$	−0.01	−13.50	16.94	−5.41	14.97	−0.37	−5.91
$GkSS_{363}$-$GkSS_{300}$	−0.89	−19.80	20.93	−3.01	23.49	−2.98	−5.23
$MjSS_{363}$-$MjSS_{300}$	2.40	−10.28	11.70	1.17	11.31	3.87	−6.59
$PhSS_{363}$-$PhSS_{300}$	−1.64	−14.24	11.74	−1.34	16.94	−3.34	−2.90

[a]Total intraprotein HB is not equal to the sum of MM, MS, and SS HBs due to the merging of HB as explained in the Materials and Methods section. [b]MM HB: main chain − main chain hydrogen bonds. [c]SS HB: side chain − side chain hydrogen bonds. [d]MS hb: Main chain − side chain hydrogen bonds. [e]chr-chr HB: hydrogen bonds between charged amino acid residues. [f]phi-phi HB: hydrogen bonds between hydrophilic amino acid residues. [g]pho-pho HB: hydrogen bonds between hydrophobic amino acid residues.

for each cutoff (4 Å−8 Å). The values are calculated by adding the differences between the unique contacts of Cα atoms at 363 K/333 K and those at 300 K . It is obvious that the number of contacts gained increases with the distance cutoff. The contact gained is larger in mesophiles compared to hyperthermophiles for all cutoffs. Flexibility attained by the mesophilic proteins at 363 K is many times more than the hyperthermophiles at the same temperature, resulting in increased Cα contact gains observed. This differential behavior is also manifested in the case of *Gk*SS, where one can observe that at 363 K, it gains more contacts than at 333 K (optimum temperature of *Gk*SS), indicating the susceptibility of *Gk*SS to 363 K similar to mesophilic counterparts. The time-dependent existence matrices are analyzed to calculate the percentage existence time of each Cα contact at 300 K and 363 K (or 333 K). The Cα contacts are classified into different bins based on their percentage existence time (Supporting Information Table S2). The short-lived contacts (% existence time ≤10%) elevated to a greater extent at 363 K in mesophilic than hyperthermophilic proteins for all cutoffs. As expected, the contact gains of $GkSS_{333}$ are analogous to hyperthermophilic gains at 363 K, and the contact gains of $GkSS_{363}$ parallel with that of the mesophilic protein at 363 K. A higher number of long-lived contacts (percent existence time 90−100%), at distant cutoffs of 5 Å−8 Å, are lost at 363 K in mesophiles than hyperthermophiles. These observations connote that the greater flexibility of mesophiles at 363 K resulted in gaining more short-lived contacts, at the same time losing more of the long-lived ones, which eventually progress at longer time scales leading to the denaturation of the protein.

Hydrophobic Contacts. In continuation with the above, the unique contacts between all the hydrophobic atoms (C, Cα, Cβ, Cδ, Cδ1, Cδ2, Cε, Cε1, Cε2, Cε3, Cγ, Cγ1, Cγ2, Cζ, Cζ2, Cζ3, Cη2) are calculated with a distance cutoff of 4 Å to 5 Å, and the total gain of unique hydrophobic contacts (net increase in contact at 333 K/363 K from those at 300 K) is calculated. Table 2 gives the summation of these contact gains of each atom. The results show that these contact gains are more in mesophiles than in hyper-/thermophiles. A classification of the unique hydrophobic contacts gains based on the percentage existence time is given in the Supporting Information Table S3. It clearly shows that momentary contacts (% existence time in the range of 0−20%) are higher, and more long-lived contacts (with % existence time >90%) are lost in mesophiles. It is established that a higher number of hydrophobic contacts (or hydrophobic core) thermally stabilizes the protein,[64] but the increase in the hydrophobic contacts as observed in the present study is due to thermal vibrations which do not lead to the formation of stable and sustained hydrophobic cores. Thus, the momentary increase of hydrophobic contact gains does not imply stability in mesophilic proteins. Although the unique hydrophobic contacts are increasing at higher temperature, the time averaged hydrophobic contacts actually reduce at a higher temperature in all proteins. The time averaged hydrophobic contacts are calculated, and a difference of this value between 363 K/333 K and 300 K is listed in Supporting Information Table S4. The loss of time averaged hydrophobic contacts are slightly more in mesophiles than in hyperthermophiles. But in *Gk*SS at 333 K, the loss is significantly less compared to all other proteins, implying that the hydrophobic interactions may play a major role in the thermal stability of *Gk*SS.

Hydrogen Bonding (HB). The HB and hydrophobic interactions are known to play important roles in the folding, stability, and functioning of the protein. According to a previous study,[65] the strength of HB in protein is ∼5 kcal/mol in isolated systems and about ∼1.5 kcal/mol in water. They have been implicated in the stability of the thermophilic proteins at high temperatures. One such study[40] shows that the intrawater HBs in the hydration shell of mesophilic proteins are comparatively weaker than that of the thermophilic proteins due to larger hydrophobic patches on the surface of the mesophilic proteins. Thus, thermophilic proteins have a continuous chain of intrawater HBs on their surface which stabilizes them at higher temperatures.[40]

The criteria used for calculating HBs are mentioned in the Materials and Methods section. The proteins from different temperature classes considered in this study have more or less a similar number of time averaged intraprotein HBs at 300 K in-line with the observations made in our earlier observation.[47] However, time averaged (4 ns−50 ns) HB calculations at 363 K (Table 3) indicate that the mesophilic proteins lose as much as ∼8−18 HBs with respect 300 K, unlike hyperthermophiles which retain almost the same number of HBs at both temperatures. Further, the effect of temperature on main chain−main chain (MM), main chain−side chain (MS), and side chain−side chain (SS) HBs are analyzed. Although the MM HBs are more numerous compared to other types, they are more fragile, and at least 10 MM HBs are broken in all proteins at higher temperatures. The MS HBs did not show any unique trend, whereas the SS HBs are comparatively stable and increase significantly in thermophilic and hyperthermophilic proteins but remain almost unchanged in mesophilic counterparts. Thus, the SS HBs are main contributors to the overall increase in the HBs of hyper-/thermophiles. The HBs are calculated among the amino acids having a similar/different chemical nature (charged−charged, hydrophilic−hydrophilic, hydrophobic−hydrophobic and between charged-other residues, hydrophilic-other residues, hydrophobic-other residues,

hydrophobic−hydrophilic, charged-hydrophilic, charged-hydrophobic residues). Among these, HBs between the charged residues elevate at 333 K/363 K comparatively more in hyper-/thermophilic than mesophilic proteins (Table 3). Except for HBs between charged−charged residues, no other group showed any unique tendency in hyper-/thermophiles concluding that charged residues are more densely populated in hyper-/thermophiles which forms increased HBs at high temperatures. In addition, the HBs among the secondary structures (α-helices and β-strands) and between these secondary structures and residues in the rest of the protein are calculated (data not shown). The HBs of α-helices are a poor discriminator, while the HBs between β-strands and the rest of the protein increased marginally in hyperthermophilic proteins at 363 K. This probably gives an explanation for the relative stability of the β-strands to temperature raise compared to α-helices.

Effect of temperature on the HBs between protein and water is given in Table 4. It gives the time averaged values of HBs at respective temperatures. It is seen that the hyperthermophilic proteins interact less with water at 363 K compared to mesophilic or thermophilic proteins due to the breakage of relatively more HBs of side chain (compared to main chain) with water at higher 363 K. Especially, the charged and hydrophobic amino acids in hyperthermophiles lose more HBs with water than thermophiles or mesophiles. At higher temperatures, the charged residues losing HB with water would form HB with other charged residues in hyperthermophilic proteins, as stated in the previous paragraph.

Table 4. Difference in the Time Averaged Values of Protein-Water HBs between 363/333 K and 300 K

	protein-water HB	M-wat HB[a]	S-wat HB[b]	chr-wat HB[c]	phi-wat HB[d]	pho-wat HB[e]
$EcSS_{363}$-$EcSS_{300}$	−80.99	−35.36	−45.63	−55.00	−17.53	−8.36
$EhSS_{363}$-$EhSS_{300}$	−41.96	−15.33	−26.63	−11.21	−25.97	−4.69
$GkSS_{333}$-$GkSS_{300}$	−58.54	−9.42	−49.10	−48.99	−1.00	−8.48
$GkSS_{363}$-$GkSS_{300}$	−114.35	−26.63	−87.70	−91.06	−6.78	−16.47
$MjSS_{363}$-$MjSS_{300}$	−103.35	−40.73	−62.60	−74.25	−17.07	−11.97
$PhSS_{363}$-$PhSS_{300}$	−104.69	−33.13	−71.57	−77.52	−8.72	−18.42

[a]M-wat HB: main chain - water hydrogen bonds. [b]S-wat HB: side chain - water hydrogen bonds. [c]chr-wat HB: charge residues - water hydrogen bonds. [d]phi-wat HB: hydrophilic residues - water hydrogen bonds. [e]pho-wat HB: hydrophobic residues - water hydrogen bonds.

The existence time of the HBs (expressed in percentage) is calculated based on the total number of times a given HB is present out of the total number of frames considered. The Supporting Information Table S5 lists the difference in the number of HBs (in each bin) between 363 K/333 K and 300 K simulations classified based on the existence time. The existence matrix of the HBs indicated that the number of unique HBs increases at higher temperature but more predominantly in mesophilic proteins mainly due to increased flexibility. At higher temperature, although the time averaged HBs reduce in mesophiles, a higher number of unique HBs are formed on account of increased flexibility. The HBs with less than or equal to 20% existence time are increased at 333 K/363 K, implying that most of the additional HBs that are formed at higher temperature are short-lived, and most of the HBs with an existence time greater than 50% (considerably long-lived) are broken at 363 K. Mesophiles, being more flexible, have a higher increase in the short-lived HBs than hyper-/thermophiles which are less flexible. Thus, the short-lived HBs may not actually contribute to the thermostability of the proteins and are generated due to the acquired flexibility at higher temperature. It is quite surprising that an equal number of long-lived HBs are broken in proteins of all temperature classes. The *Gk*SS (which is more stable at 333 K than 363 K) has more short-lived HBs gained at 363 K than 333 K and lose lesser long-lived HBs at 333 K than at 363 K correlating well with the mesophilic proteins. Although the present observation based on HBs existence time do not ascertain its contribution to the thermostability, it clearly establishes the effect of flexibility on HBs that enables one to delineate between mesophiles and hyper-/thermophiles proteins.

Salt-Bridges (SB). One of the most important noncovalent interactions that are known to stabilize the hyper-/thermophilic proteins are the ion-pair interactions or salt-bridges (SB). They are electrostatic interactions between oppositely charged residues in protein. In most of the hyper-/thermophilic

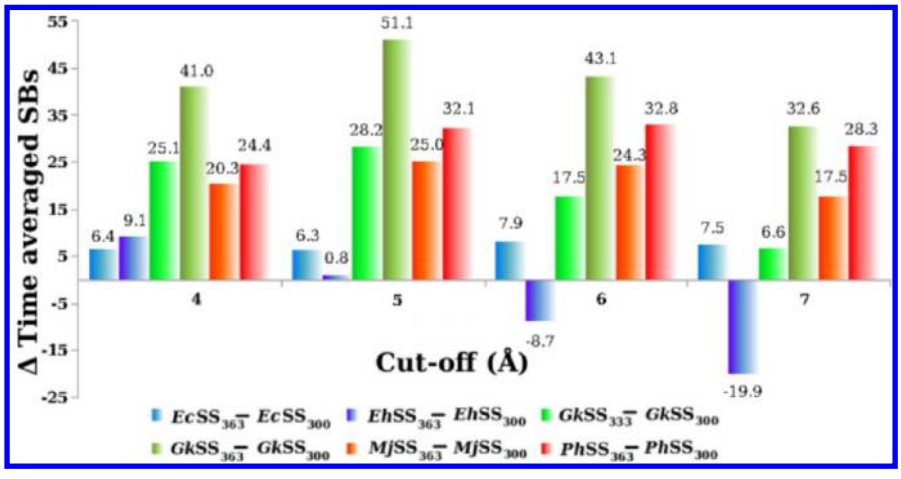


Figure 6. A bar graph showing the difference in time averaged SBs between 363 K/33 3K and 300 K at different distance cutoff with the corresponding values indicated above.

dx.doi.org/10.1021/ci400306m | J. Chem. Inf. Model. 2013, 53, 2448−2461

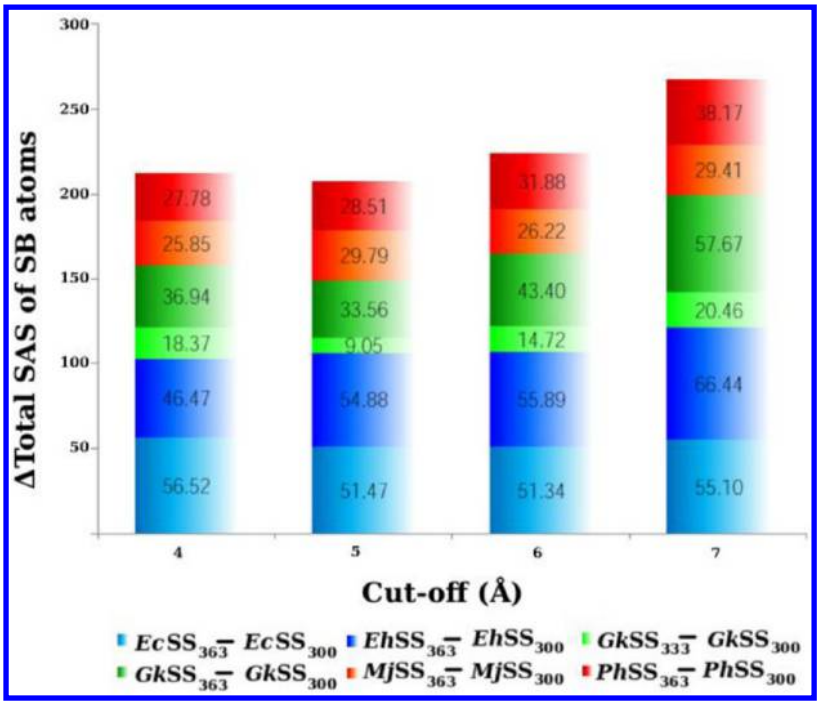


Figure 7. A stacked bar graph showing the difference in total solvent accessible surface area (sasa) of all SBs between 363 K/333 K and 300 K (corresponding values are given inside each bar).

proteins, these charged residues are abundant compared to mesophilic counterparts implying that it may have a relatively higher number of SBs. Formation of a SB requires desolvation of the charged residues and subsequently forming electrostatic interactions with the oppositely charged residues. When the gain from electrostatic interaction is more than the loss due to desolvation of the exposed charged residues, the ionic interactions are stabilized. At high temperature, due to the lowering of the dielectric constant of water, the loss due to the desolvation of the charged residues will be minimized, thus, encouraging the formation of SBs. In the present study, the time-averaged values, accessibility, and the number of unique SBs with its existence time are calculated over the trajectory and classified at different cutoffs (4 Å to 7 Å).

The time averaged values of the number of SBs at different cutoff for all simulations are illustrated in Figure 6. It can be seen that the number of SBs increases to a greater extent (at all cutoff) in hyper-/thermophiles than mesophiles at 363 K, and it may even reduce in the case of mesophilic proteins, as observed in *Eh*SS. The measure of increase is at least two times higher in hyper-/thermophiles, emphasizing the role of SBs to the thermal stability. At 7 Å, even in hyper-/thermophiles, the increase of long-range SBs at 363 K is lower compared to other distance cutoffs, indicating that long-range electrostatic interactions beyond 7 Å might not provide as much thermal stability as the short-range interactions.

The unique SBs (in both chains) are calculated by considering those formed between the same two residues but different atoms as a single SB. Supporting Information Table S6 lists the number of unique SBs between the residues at different distance cutoffs classified based on the percentage existence time. The differences in the number of unique SBs between two temperatures are more or less similar in proteins of all temperature classes and at all cutoffs. Though, it can be noticed in Table S6 (highlighted in yellow), even at 300 K, hyperthermophiles have a significantly higher number of unique SBs compared to mesophilic proteins mainly at longer cutoffs, but stronger SBs (at small cutoffs) i.e. those with a longer existence time are comparatively lesser in hyperthermophiles, which implies that hyperthermophiles have relatively weaker SBs at 300 K. At 363 K, both the short- and long-lived SBs are mostly higher in hyperthermophiles except at 4 Å cutoff.

The sum of surface accessibility of atoms forming SBs is calculated at both temperatures (300 K, 333 K/363 K). Surprisingly, at 363 K, the total accessibility of all the SBs did not appear to be more in hyper-/thermophiles compared to mesophiles, unlike the general opinion that a higher number of SBs are located on the protein surface in hyperthermophiles. However, the difference between the total accessibility of all SBs at 363 K/333 K and at 300 K (Figure 7) indicates that (at all cutoffs) the hyperthermophiles (and thermophile at 333 K) have a lesser value than mesophiles. An explanation for this could be attributed to the opening up of mesophilic protein eventually exposing residues to water thus increasing the accessibility to a greater extent than hyper-/thermophiles. The general conclusion is that the surface accessibility of SBs increases as the temperature rises, but due to the instability of mesophilic proteins, their accessibility increases much higher than that of thermo-/hyperthermophiles. The present study does not support the hypothesis that hyper-/thermophiles have a higher number of SBs located on the surface of protein or they are more accessible.

CONCLUSION

In the present study, an attempt is made to delineate the mesophilic (*E. coli*, *E. chaffeensis*), thermophilic (*G. kaustophilus*), and hyperthermophilic (*M. jannaschii*, *P. horikoshii*) proteins on the basis of their thermal stability by studying their dynamic aspects at 300 K and 333 K/363 K using molecular dynamics. The main focus is to identify the features that can be used to differentiate proteins from these diverse

temperature classes during the initial stages of denaturation of the least stable protein. Both mesophilic and thermophilic proteins show similar signs of instability at 363 K, while thermophilic and hyperthermophilic proteins maintain their integrity at 333 K and 363 K, respectively. At 363 K, both mesophilic (to a greater extent) and thermophilic (to a lesser extent) proteins display a considerable increase in their radius of gyration indicating a gradual release of the structural network. The deviation from the initial structure is more prominent in these two temperature classes. The regions of the protein near the active site and at the dimeric interface fluctuate more in mesophiles and thermophiles. The mean square displacement curves proved to be the best approach to distinguish protein belonging to distinct temperature classes and it also establishes the Somero's corresponding states hypothesis. A change in the composition of the secondary structure (α-helix+β-sheet+β-bridge+turn) is evident in mesophiles with the α-helix being more sensitive to temperature than the β-sheet at 363 K although the overall increase in time-averaged sasa at 363 K is not distinguishable enough between the temperature classes and the residues get exposed to a greater extent in mesophiles at 363 K. Further, hyperthermophilic and thermophilic proteins have more atoms buried (at 363 K and 333 K, respectively) than mesophilic proteins which provides thermal stability to their structures through hydrophobic interactions.

At 363 K, mesophiles show drastic variations in the intra-Cα distances which are more prominent at a long-range of $\geq$1.5 nm unlike hyperthermophiles or thermophiles (at 333 K). Owing to the increased flexibility in mesophilic proteins at 363 K, comparatively a higher number of contacts (Cα at cutoff of 4 Å–8 Å and hydrophobic at cutoff 4 Å–5 Å) are acquired. Although the mesophiles acquire a higher number of hydrophobic contacts, it does not correlate with the thermophilicity of the proteins. At 363 K, relative to mesophilic proteins, a higher number of intraprotein HBs are formed, and a larger number of protein-water HBs are broken in hyperthermophiles. Thus, unlike the structural analysis, the results of the dynamics support the involvement of HBs to thermal stability. The present study also emphasizes the contribution of SBs in thermal stability of hyper-/thermophiles which show a 2-fold increase of the average SB interactions at 363 K compared to mesophilic proteins. We hope that the present investigation would provide a basic understanding of how the hyper-/thermophilic enzymes manages to attain thermal stability in spite of being more similar (in sequence and structure) to mesophilic counterparts. In addition, this would lay a foundation for the future MD studies of this enzyme with the substrates to elucidate how the catalysis, involving temperature sensitive substrates, takes place at such high temperature.

■ ASSOCIATED CONTENT

Supporting Information

Tables describing the time dependent properties. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +91-80-22933059/22933060. Fax: +91-80-23600683. E-mail: sekar@physics.iisc.ernet.in, sekar@serc.iisc.ernet.in.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the facilities offered by the Interactive graphics and the Supercomputer Education and Research Centre. The authors thank the Department of Science and Technology (DST) for financial support.

■ REFERENCES

(1) Stetter, K. O. Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. London, Ser. B* **2006**, *361* (1474), 1837–1843.
(2) Imanaka, T. Molecular bases of thermophily in hyperthermophiles. *Proc. Jpn. Acad., Ser. B* **2011**, *87* (9), 587–602.
(3) Perutz, M. F.; Raidt, H. Stereochemical basis of heat stability in bacterial ferredoxins and in hemoglobin A2. *Nature* **1975**, *255* (5505), 256–259.
(4) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 1–43.
(5) Wright, H. T. Nonenzymatic deamidation of asparaginyl and glutamyl residues in proteins. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26* (1), 1–52.
(6) Zale, S. E.; Klibanov, A. M. Why does ribonuclease irreversibly inactivate at high temperatures? *Biochemistry* **1986**, *25* (19), 5432–5444.
(7) Jaenicke, R.; Böhm, G. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **1998**, *8* (6), 738–748.
(8) Haney, P. J.; Badger, J. H.; Buldak, G. L.; Reich, C. I.; Woese, C. R.; Olsen, G. J. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (7), 3578–3583.
(9) Argos, P.; Rossmann, M. G.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. Thermal stability and protein structure. *Biochemistry* **1979**, *18* (25), 5698–5703.
(10) Facchiano, A. M.; Colonna, G.; Ragone, R. Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Eng.* **1998**, *11* (9), 753–760.
(11) Choi, I.-G.; Bang, W.-G.; Kim, S.-H.; Yu, Y. G. Extremely thermostable serine-type protease from Aquifex pyrophilus. Molecular cloning, expression, and characterization. *J. Biol. Chem.* **1999**, *274* (2), 881–888.
(12) Thompson, M. J.; Eisenberg, D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J. Mol. Biol.* **1999**, *290* (2), 595–604.
(13) Beeby, M.; O'Connor, B. D.; Ryttersgaard, C.; Boutz, D. R.; Perry, L. J.; Yeates, T. O. The genomics of disulfide bonding and protein stabilization in thermophiles. *PloS Biol.* **2005**, *3* (9), 1549–1558.
(14) Pace, C. N. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **1992**, *226* (1), 29–35.
(15) Goldstein, R. A. Amino-acid interactions in psychrophiles, mesophiles, thermophiles and hyperthermophiles: Insights from the quasi-chemical approximations. *Protein Sci.* **2007**, *16* (9), 1887–1895.
(16) Dong, H.; Mukaiyama, A.; Tadokoro, T.; Koga, Y.; Takano, K.; Kanaya, S. Hydrophobic effect on the stability and folding of a hyperthermophilic protein. *J. Mol. Biol.* **2008**, *378* (1), 264–272.
(17) Paiardini, A.; Sali, R.; Bossa, F.; Pascarella, S. "Hot cores" in proteins: comparative analysis of the polar contact area in structures from hyper/thermophilic and mesophilic organisms. *BMC Struct. Biol.* **2008**, *8*, 14.
(18) Kannan, N.; Vishveshwara, S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13* (11), 753–761.
(19) Ibrahim, B. S.; Pattabhi, V. Role of weak interactions in thermal stability of proteins. *Biochem. Biophys. Res. Commun.* **2004**, *325* (3), 1082–1089.

(20) Vetriani, C.; Maeder, D. L.; Tolliday, N.; Yip, K. S.; Stillman, T. J.; Britton, K. L.; Rice, D. W.; Klump, H. H.; Robb, F. T. Protein thermostability above 100 degreeC: key role for ionic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (21), 12300−12305.
(21) Xiao, L.; Honig, B. Electrostatic contribution to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **1999**, *289* (5), 1435−1444.
(22) Kumar, S.; Tsai, C.-J.; Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **2000**, *13* (3), 179−191.
(23) Szilàgyi, A.; Závodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **2000**, *8* (5), 493−504.
(24) Karshikoff, A.; Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: "traffic rule" for hot roads. *Trends Biochem. Sci.* **2001**, *26* (9), 550−556.
(25) Vogt, G.; Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Struct. Fold. Des.* **1997**, *2* (4), S40−S46.
(26) Vogt, G.; Woell, S.; Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **1997**, *269* (4), 631−643.
(27) Zhou, H.-X. Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophys. J.* **2002**, *83* (6), 3126−3133.
(28) Dougherty, D. Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr and Typ. *Science* **1996**, *271* (5246), 163−168.
(29) Gromiha, M. M.; Thomas, S.; Santosh, C. Role of cation-pi interactions to the stability of thermophilic proteins. *Prep. Biochem. Biotehnol.* **2002**, *32* (4), 355−62.
(30) Berezovsky, I. N.; Shakhnovich, E. I. Physics and evolution of thermophiles adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (36), 12742−12747.
(31) Radestock, S.; Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins* **2011**, *79* (4), 1089−1108.
(32) Thoma, R.; Henning, M.; Sterner, R.; Kirschner, K. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermatoga maritima*. *Structure* **2000**, *8* (3), 265−276.
(33) Sawle, L.; Ghosh, K. How do thermophilic proteins and proteomes withstand high temperature. *Biophys. J.* **2011**, *101* (1), 217−227.
(34) Nojima, H.; Ikai, A.; Oshima, T.; Noda, H. Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J. Mol. Biol.* **1977**, *116* (3), 429−442.
(35) Cooper, A. Protein heat capacity: an anomaly that maybe never was. *J. Phys. Chem. Lett.* **2010**, *1* (22), 3298−3304.
(36) Sterpone, F.; Melchionna, S. Thermophilic proteins: insight and perspective from *in silico* experiments. *Chem. Soc. Rev.* **2012**, *41* (5), 1665−1676.
(37) Prabhu, N. V.; Sharp, K. A. Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521−548.
(38) Vijayakumar, S.; Vishveshwara, S.; Ravishanker, G.; Beveridge, D. L. Differential stability of beta-sheets and alpha-helices in beta-lactamase: a high temperature molecular dynamics study of unfolding intermediates. *Biophys. J.* **1993**, *65* (6), 2304−2312.
(39) Purmonen, M.; Valjakka, J.; Takkinen, K.; Laitinen, T.; Rouvinen, J. Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Eng., Des. Sel.* **2007**, *20* (11), 551−559.
(40) Sterpone, F.; Bertonati, C.; Briganti, G.; Melchionna, S. Key role of proximal water in regulating thermostable proteins. *J. Phys. Chem. B* **2009**, *113* (1), 131−137.
(41) Pizzitutti, F.; Marchi, M.; Sterpone, F.; Rossky, P. J. How protein surface induce anomalous dynamics of hydration water. *J. Phys. Chem. B* **2007**, *111* (26), 7584−7590.
(42) Bleicher, L.; Prates, E. T.; Gomes, T. C.; Silveira, R. L.; Nascimento, A. S.; Rojas, A. L.; Golubev, A.; Martínez, L.; Skaf, M. S.; Polikarpov, I. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from Rhodothermus marinus and molecular dynamics simulations. *J. Phys. Chem. B* **2011**, *115* (24), 7940−7949.
(43) Somero, G. N. Temperature adaptation of enzymes: biological optimization through structure-function compromises. *Annu. Rev. Ecol. Syst.* **1978**, *9*, 1−29.
(44) Radestock, S.; Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins* **2011**, *79* (4), 1089−1108.
(45) Levdikov, V. M.; Barynin, V. V.; Grebenko, A. I.; Melik-Adamyan, W. R.; Lamzin, V. S.; Wilson, K. S. The structure of SAICAR synthetase: an enzyme in the *de novo* pathway of purine nucleotide biosynthesis. *Structure* **1998**, *6* (3), 363−376.
(46) Ginder, N. D.; Binkowski, D. J.; Fromm, H. J.; Honzatko, R. B. Nucleotide complexes of Escherichia coli phosphoribosylaminoimidazole succinocarboxamide synthetase. *J. Biol. Chem.* **2006**, *281* (30), 20680−20688.
(47) Manjunath, K.; Kanaujia, S. P.; Kanagaraj, S.; Jeyakanthan, J.; Sekar, K. Structure of SAICAR synthetase from Pyrococcus horikoshii OT3: Insights into thermal stability. *Int. J. Biol. Macromol.* **2013**, *53*, 7−19.
(48) Zhang, R.; Skarina, T.; Evdokimova, E.; Edwards, A.; Savchenko, A.; Laskowski, R.; Cuff, M. E.; Joachimiak, A. Structure of SAICAR synthase from Thermotoga maritima at 2.2 angstroms reveals an unusual covalent dimer. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2006**, *62*, 335−339.
(49) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22* (2), 195−201.
(50) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features an development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(51) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435−447.
(52) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimization for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657−1666.
(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(54) Darden, T.; York, D.; Pederesen, L. Particle mesh Ewald: an N.log(N) method for sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(55) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.
(56) Nose, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **1984**, *81* (1), 511−519.
(57) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31* (3), 1695−1697.
(58) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **1981**, *52* (12), 7182−7190.
(59) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23* (21), 2947−2948.
(60) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **2006**, *64* (3), 559−574.
(61) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577−2637.
(62) Monera, O. D.; Sereda, T. J.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. Relationship of sidechain hydrophobicity and α-helical propensity on the stability of single-stranded amphipathic α-helix. *J. Pept. Sci.* **1995**, *1* (5), 319−329.

(63) Marcos, E.; Jiménez, A.; Crehuet, Ramon. Dynamic fingerprints of protein thermostability revealed by long molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1129−1142.
(64) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Shirley, B. A.; Hendricks, M. M.; Iimura, S.; Gajiwala, K.; Scholtz, J. M.; Grimsley, G. R. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* **2011**, *408* (3), 514−528.
(65) Sheu, S. Y.; Yang, D. Y.; Selzle, H. L.; Schlag, E. W. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (22), 12683−12687.