JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

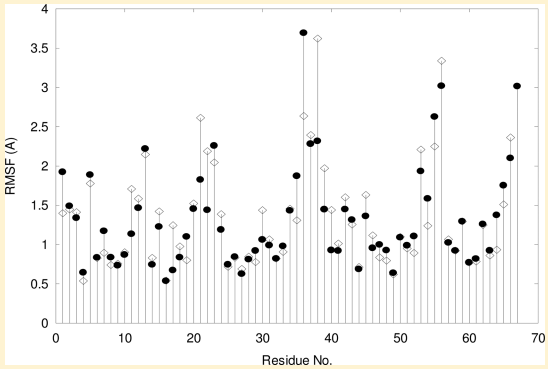
Do Homologous Thermophilic−Mesophilic Proteins Exhibit Similar Structures and Dynamics at Optimal Growth Temperatures? A Molecular Dynamics Simulation Study

Sohini Basu and Srikanta Sen*

Molecular modeling Section, Biolab, Chembiotek, TCG Lifesciences Ltd., Bengal Intelligent Park, Tower-B 2nd Floor, Block-EP & GP, Sector-V, Salt Lake Electronic Complex, Calcutta-700091, India

Supporting Information

ABSTRACT: Structure and dynamics both are known to be important for the activity of a protein. A fundamental question is whether a thermophilic protein and its mesophilic homologue exhibit similar dynamics at their respective optimal growth temperatures. We have addressed this question by performing molecular dynamics (MD) simulations of a natural mesophilic−thermophilic homologue pair at their respective optimal growth temperatures to compare their structural, dynamical, and solvent properties. The MD simulations were done in explicit aqueous solvent under periodic boundary and constant pressure and temperature (CPT) conditions and continued for 10.0 ns using the same protocol for the two proteins, excepting the temperatures. The trajectories were analyzed to compare the properties of the two proteins. Results indicated that the dynamical behaviors of the two proteins at the respective optimal growth temperatures were remarkably similar. For the common residues in the thermophilic protein, the rms fluctuations have a general trend to be slightly higher compared to that in the mesophilic counterpart. Lindemann parameter values indicated that only a few residues exhibited solid-like dynamics while the protein as a whole appeared as a molten globule in each case. Interestingly, the water−water interaction was found to be strikingly similar in spite of the difference in temperatures while, the protein−water interaction was significantly different in the two simulations.



INTRODUCTION

Common proteins are denatured and become nonfunctional when the temperature is increased beyond 50 °C. However, in hot places like hot springs and hot vents at the bottom of the sea, organisms have been found that are capable of growing and functioning properly at high temperatures.[1,2] Such organisms are called thermophiles. If the optimal growth temperature (T_{opt}) of the organism is in the range 50 < T_{opt} < 80 °C, it is called a moderate thermophile. Organisms with T_{opt} > 80 °C are called hyperthermophiles. Ordinary organisms have their optimal growth temperatures in the range 20−50 °C and are known as mesophilic organisms. Proteins isolated from thermophilic or hyperthermophilic organisms remain structurally stable and functionally active at much higher temperatures where their mesophilic counterparts are denatured and become inactive.[3,4] Thermophilic or hyperthermophilic proteins and their mesophilic counterparts generally share a high degree of similarity in their amino acid sequences and hence in their three-dimensional structures.[5,6] It is of great interest to study the proteins that are stable at high temperatures not only for their unusual stability but also because of their potential industrial uses.[1,7−9] Understanding the physical basis of the properties of thermophilic proteins is expected to be useful in designing tailor-made thermostable proteins. There are already a few approaches available for designing thermostable proteins.[10−18]

Extensive efforts have been devoted in elucidating the physicochemical origin of thermostability, and such studies have suggested that not a single mechanism but a variety of factors can be responsible for this enhanced stability. Such factors include optimized electrostatic interactions like increased number of salt-bridges and H-bonds, improved packing, networks of hydrogen bonds, increased hydrophobic interactions, decreased number and volume of internal cavities, etc.[19−29] Until now, no general rules for determining thermostability are found, and different thermophilic proteins appear to achieve their increased thermostabity due to different combinations of a few of the above-mentioned factors. Comparative studies have further pointed out that many thermophilic proteins have increased numbers of salt bridges compared to their homologous mesophilic counterparts.[19,23,30−34] Moreover, a large number of salt bridges are generally found to be present on the surface of thermophilic proteins, and these salt bridges are expected to participate not only in intraprotein interactions but also in interdomain and

Received: October 4, 2012

   dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

intersubunit interfaces of proteins for their further stabilization.[35,36]

It is now well-accepted that the function of a protein is not only strongly dependent on its 3D structure but also on its dynamics and it is generally believed that the thermophilic proteins at its optimal growth temperature exhibit dynamics similar to that of its mesophilic counterpart at the respective optimal growth temperature.

It is now well established that molecular dynamics simulations are quite useful in investigating the structural and dynamical properties of biomolecular systems at the atomic level. A considerable amount of work in understanding the dynamics and stability of proteins has been done based on experimental, theoretical, and molecular dynamics (MD) simulation approaches.[37−44] The review by Sterpone and Melchionna provides a good summary of such activities.[45] The findings of our present study are mostly consistent with the results of the previous studies. For example, previous studies showed systematically larger amplitudes of fluctuations with the degree of thermostability,[37−45] and we also have observed the same feature. MD simulations have been used to assess the relative stability of homologue proteins. This is achieved by comparing the protein behavior at high temperatures. A critical validation of simulations relies on the observation that thermophiles resist more to high temperatures than the mesophilic homologues. A recent molecular dynamics simulation based report demonstrates that at a lower temperature the structure of a thermophilic protein is considerably more rigid than that of its mesophilic homologue.[44] This finding again goes favorably with our current MD simulation results. Moreover, this is consistent with the fact that most hyperthermophilic proteins are found to be poorly active or inactive at room temperature.[46,47] However, most importantly, the fundamental issue whether the thermophilic and its mesophilic homologue exhibit similar dynamics at their respective optimal growth temperatures has not been discussed explicitly in the previous MD simulation based works and has not yet been resolved completely. Being motivated by this fact, in the present work we have given an effort to explicitly investigate this important issue. We have therefore performed MD simulations of a naturally occurring thermophilic−mesophilic protein pair to compare their structural, dynamical, and thermostability properties at their respective optimal growth temperatures.

There is another interesting issue which is worth detailed investigations. It is known that there are differences in the amino acid sequences and contents between a mesophilic protein and its thermophilic counterpart, and these differences are mainly responsible for their difference in stabilities. However, at the two different optimal growth temperatures the same normal water is present as the solvent. Thus, it appears to be quite interesting to investigate also the solvent properties and the solvent−protein interactions of the mesophilic and thermophilic protein pair at their optimal growth temperatures from the MD simulation trajectories. With these intentions in mind, we have chosen a natural homologous mesophilic−thermophilic protein pair whose crystal structures as well as the optimal growth temperatures are known. We have performed separate molecular dynamics simulations for the mesophilic and thermophilic proteins of the pair at their respective optimal growth temperatures under CPT condition in explicit water molecules as solvent with periodic boundary conditions. In each case a trajectory of 10 ns has been produced and analyzed. For clarity and simplicity, we represent the optimal growth temperatures of the mesophilic protein and the themophilic proteins as T_{opt}^{Meso} and T_{opt}^{Thermo}, respectively. We have also performed 10 ns MD simulations of the mesophilic protein at T_{opt}^{Thermo} and the vice versa as control experiments. Detailed analyses of the trajectories have been summarized in this paper.

MATERIALS AND METHODS

(A) Choice of Mesophilic−Thermophilic Protein Pair. We have chosen the mesophilic cold shock protein from *Bacillus subtilis* (*Bs*-CspB),[48] and the thermophilic cold shock protein from *Bacillus caldolyticus* (*Bc*-Csp),[49] as a representative of natural mesophilic−thermophilic protein pair because their crystal structures (1csp, 1c9o) as well as the respective optimal growth temperatures (T_{opt}^{Meso} = 35.0 °C and T_{opt}^{Thermo} = 72.0 °C) are known.[50,51] Each of the two structures contains a few loops and a central beta-barrel motif as the main ordered structural unit. This protein pair is also experimentally well studied.

(B) Molecular Dynamics Simulation Protocol. We have followed standard methods of molecular dynamics simulations in explicit water molecules as solvent. The protein structures were prepared first by adding H-atoms to the crystal structures using CHARMM H-build facility. The resulting structures were then primarily refined by performing energy minimization in vacuum by 1000 steepest descent (SD) steps using a distance dependent dielectric constant in order to remove any bad steric conflicts in the crystal structures. Each of these energy minimized structures were then solvated by placing each of them separately in a box of dimension 40.0 Å × 40.0 Å × 45.0 Å containing 2405 pre-equilibrated Tip3P water molecules.[52] Any water molecule that had an atom closer than 2.8 Å to any non-H atom of the protein was removed. Thus, the solvated thermophilic protein consisted of 2043 water molecules, and the mesophilic counterpart contained 2024 water molecules. In the case of the mesophilic protein, the number of water molecules is less as the mesophilic protein consists of one residue more and is bigger than the thermophilic homologue. In order to make each system electrostatically neutral, we added two sodium ions to the solvated thermophilic system and six sodium ions to the solvated mesophilic system. The ions were placed by replacing the energetically most unfavorable water molecules in the respective systems. Each of these two solvated proteins was then refined by energy minimization by 500 SD steps under periodic boundary condition keeping the protein fixed. This process allows only the water molecules to adjust their positions and orientations around the protein in the box in the most favorable way. Each of the systems were then finally energy minimized by 500 SD steps using boundary condition without any constraint. These solvated systems were then used for performing separate molecular dynamics simulations at the respective desired temperatures.

The MD simulation of this solvated system was executed following CPT method using leapfrog algorithm under periodic boundary conditions. The Shake algorithm[53] was used throughout the simulation to constrain all the bonds involving H-atoms. A time step of 0.002 ps was used for integrating the Newton's equations of motion. First, for the mesophilic protein, the solvated system was heated from 108 to 308 K during the first 10 ps by assigning velocities to the atoms every 50 steps so that the system reaches a value close to the target optimal growth temperature of 308 K. Then, the simulation was continued for equilibration at a temperature 308 K by velocity

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

assignment for 20 ps followed by equilibration at 308 K temperature by velocity scaling for 9970 ps. During this period the temperature was kept constant by scaling atomic velocities only if the temperature were outside the window of 308 ± 5 K. The nonbonded interactions were computed using a spherical cutoff 12.0 Å, and the interaction was switched smoothly to zero over the range of 10.0−11.0 Å. The nonbonded pair list was updated every 20 steps. We have not used Ewald method for the long-range interactions as it has been shown to introduce artificial order in the structure.[54] The spherical cutoff method has produced realistic trajectories.[55−57] The dielectric constant value used was 1.0. The overall simulation was done under the CPT condition at a temperature of 308 K and pressure of 1 atm. The trajectories were saved every 200 steps. The last 4 ns of the trajectory was considered for computing various average properties of the system. A similar MD simulation was performed with the solvated thermophilic protein using the same protocol excepting the temperature that was 345 K (T_{opt}^{Thermo}).

As control experiments we have also performed 10 ns MD simulations of the mesophilic protein at T_{opt}^{Thermo} and vice versa following the same protocol.

All the MD simulations were performed by employing the program package CHARMM, version 26, and its parameter set.[58,59]

(C) Analysis. *Structure, Stability, and Dynamics of Proteins.* The structural difference between two conformers of a protein is expressed in terms of root mean squared deviation (RMSD). The time dependent RMSD with reference to its initial ($t = 0$) structure is defined as

$$\mathrm{RMSD}(t) = \left[\frac{1}{N}\sum_i (\Delta r_i^2)\right]^{1/2} \tag{1}$$

where, $\Delta r_i^2 = (r_i^t - r_i^{ini})^2$ and r_i^t is the position of the ith atom at the tth time frame and r_i^{ini} is the initial position of the ith atom. Thus, RMSD(t) is a single measure to represent the overall deviation of the structure at the tth time frame with reference to the initial structure. N is the total number of atoms in the protein.

The Lindemann parameter is a measure of the disorder in a system.[60−64] The Lindemann criterion is a useful way to characterize the dynamical behavior of a protein in terms of "solid-like" or "liquid-like" properties. It compares the atomic fluctuation amplitude $\langle \Delta r_i^2 \rangle^{1/2}$ with the lattice constant a' of a crystal.[60−64] The Lindemann criterion was introduced for determining whether an infinite system is solid-like or liquid-like.[61,62] It was used also for finite systems, such as Lennard-Jones clusters and isolated homopolymers.[60,64] Lindemann parameter was used earlier to show that the interior of native proteins is solid-like, while their surface is liquid-like and the entire protein becomes solid-like at low temperature (220 K) making the protein inactive.[60] Lindemann parameter for a protein is defined as a disorder parameter (Δ_L) as given below.

$$\Delta_L = \frac{1}{a'}\left[\frac{1}{N}\sum_i (\Delta r_i^2)\right]^{1/2} \tag{2}$$

where, $\Delta r_i^2 = (r_i - \langle r_i \rangle)^2$, r_i is the position of the ith atom and $\langle r_i \rangle$ represents the configurational average and a' is a parameter equivalent to the lattice parameter. We have used a value of 4.5 for a' as suggested by Zhou and Karplus.[60] Since the maximum of the distribution of nonbonded atoms in several proteins at 300 K was found to be between 4 and 5 Å, the value of a' was chosen as 4.5 Å by Zhou and Karplus.[60]

We define the Lindemann parameter for an individual residue as

$$\Delta_L^J = \frac{1}{a'}\left[\frac{1}{N_j}\sum_i (\Delta r_{ij}^2)\right]^{1/2} \tag{3}$$

where, $\Delta r_{ij}^2 = (r_{ij} - \langle r_{ij} \rangle)^2$, r_{ij} is the position of the ith atom of the jth residue and N_j is the number of atoms in the jth residue.

The Lindemann criterion tells that particles with $\Delta_L^J < 0.15$ Å exhibit solid-like dynamics while particles with $\Delta_L^J > 0.15$ Å exhibit liquid-like dynamics.[61−64]

Root mean squared fluctuations for the residues were computed as follows.

$$\mathrm{rmsf}_j = \left[\frac{1}{N_j}\sum_i (\Delta r_{ij}^2)\right]^{1/2} \tag{4}$$

Δr_{ij}^2 and N_j have the same meaning as before.

For further comparison of the overall rms fluctuations of the trajectories of the two proteins, we defined the overall fluctuations ($F_{overall}$) as

$$F_{overall} = \left[\frac{1}{n}\sum_{i=1}^{n} \mathrm{rmsf}_i^2\right]^{1/2} \tag{5}$$

where, rmsf_i represents the fluctuation of the ith residue and n is the number of residues present in the protein. Thus, $F_{overall}$ is a single valued measure of the overall RMSF of the protein.

The stability of a protein at a given temperature is determined as the difference in the free energies (G) of the protein in its folded and unfolded states in solution and is represented as Stability = G(folded) − G(unfolded) = $\Delta H - T\Delta S$, where ΔH represents the difference in enthalpy between the folded and unfolded states, T is the absolute temperature, and ΔS represents the change in entropy between the two states. Thus, the difference (ΔStability) between the stabilities of the homologous thermophilic and mesophilic proteins is given by

$$\begin{aligned}\Delta\mathrm{Stability} &= \Delta G_{Thermo} - \Delta G_{Meso} \\ &= [\Delta H_{Thermo} - \Delta H_{Meso}] \\ &\quad - [T_{Thermo}\Delta S_{Thermo} - T_{Meso}\Delta S_{Meso}]\end{aligned}$$

In the cases of homologous protein pairs, it is unlikely that the difference in the entropic terms will be highly significant; hence for simplicity, we may ignore the difference in the overall entropic contributions, and we get $\Delta\mathrm{Stability} \approx \Delta H_{Thermo} - \Delta H_{Meso}$. The difference in enthalpy, in general, has several components coming from protein's self-energies, solvent−protein interaction energies, and solvent−solvent interaction energies and can be expressed as

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

```
CSP             MLEGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQA
C9O             MQRGKVKWFNNEKGYGFIEVEGGSDVFVHFTAIQGEGFKTLEEGQEVSFEIVQGNRGPQA
                * .*******.***:******* .******:************** ******:*******

CSP             ANVTKEA
C9O             ANVVKL-
                ***.*
```

Figure 1. Aligned amino acid sequences of the thermophilic and mesophilic proteins pair are shown. The 12 mismatched residues are highlighted in bold.

$$\begin{aligned}\Delta H_{\text{Thermo}} = {} & [E_{\text{Self}}^{\text{Thermo}}(\text{folded}) - E_{\text{Self}}^{\text{Thermo}}(\text{unfolded})] \\ & + [E_{\text{solvent-protein}}^{\text{Thermo}}(\text{folded}) \\ & - E_{\text{solvent-protein}}^{\text{Thermo}}(\text{unfolded})] \\ & + [E_{\text{solvent-solvent}}^{\text{Thermo}}(\text{folded}) \\ & - E_{\text{solvent-solvent}}^{\text{Thermo}}(\text{unfolded})] \end{aligned} \tag{6}$$

$$\begin{aligned}\Delta H_{\text{Meso}} = {} & [E_{\text{Self}}^{\text{Meso}}(\text{folded}) - E_{\text{Self}}^{\text{Meso}}(\text{unfolded})] \\ & + [E_{\text{solvent-protein}}^{\text{Meso}}(\text{folded}) \\ & - E_{\text{solvent-protein}}^{\text{Meso}}(\text{unfolded})] \\ & + [E_{\text{solvent-solvent}}^{\text{Meso}}(\text{folded}) \\ & - E_{\text{solvent-solvent}}^{\text{Meso}}(\text{unfolded})] \end{aligned} \tag{7}$$

The terms are self-explanatory. Thus,

$$\begin{aligned}\Delta\text{Stability} \approx {} & [E_{\text{Self}}^{\text{Thermo}}(\text{folded}) - E_{\text{Self}}^{\text{Meso}}(\text{folded})] \\ & - [E_{\text{Self}}^{\text{Thermo}}(\text{unfolded}) - E_{\text{Self}}^{\text{Meso}}(\text{unfolded})] \\ & + [E_{\text{solvent-protein}}^{\text{Thermo}}(\text{folded}) \\ & - E_{\text{solvent-protein}}^{\text{Meso}}(\text{folded})] \\ & - [E_{\text{solvent-protein}}^{\text{Thermo}}(\text{unfolded}) \\ & - E_{\text{solvent-protein}}^{\text{Meso}}(\text{unfolded})] \\ & + [E_{\text{solvent-solvent}}^{\text{Thermo}}(\text{folded}) \\ & - E_{\text{solvent-solvent}}^{\text{Meso}}(\text{folded})] \\ & - [E_{\text{solvent-solvent}}^{\text{Thermo}}(\text{unfolded}) \\ & - E_{\text{solvent-solvent}}^{\text{Meso}}(\text{unfolded})] \end{aligned} \tag{8}$$

The self-energy (E_{Self}) of a protein in a given 3D structure represents the total energy content of the structure including all the bonded and nonbonded interaction energy components of the protein.

The thermophilic–mesophilic protein pair has only a few mismatches in their sequences making them largely similar to each other, and it is expected that the self-energies of proteins in their unfolded states are very similar. Similarly, the overall difference in the protein–solvent interactions, considering both the folded and unfolded proteins together, is not expected to be considerable. This is because if there is a considerable difference in protein–solvent interaction in the folded states of the thermophilic and mesophilic protein, similar difference will also be there in protein–solvent interactions between their unfolded states and thus the overall contribution to stability difference will be negligible (relation 8). Likewise, the solvent–solvent interactions in the folded and unfolded states are expected to be similar. Thus, the ΔStability may be approximated as

$$\Delta\text{Stability} \approx E_{\text{Self}}^{\text{Thermo}}(\text{folded}) - E_{\text{Self}}^{\text{Meso}}(\text{folded}) \tag{9}$$

Hence, the difference ($\Delta E_{\text{Self}} = E_{\text{Self}}^{\text{Thermo}}(\text{folded}) - E_{\text{Self}}^{\text{Meso}}(\text{folded})$) in the self-energies of the thermophilic ($E_{\text{Self}}^{\text{Thermo}}(\text{folded})$) and mesophilic variant ($E_{\text{Self}}^{\text{Meso}}(\text{folded})$) can be used as an approximate measure of the stability differences between the two, where at least a major component of the stability factors is taken into account. As defined here, a negative value of ΔE_{Self} implies that the mutated thermophilic protein is more stable than the corresponding original one. The self-energy of a given conformer of the protein was calculated by CHARMM using its standard force field.

SASA Calculation. The solvent accessible surface areas (SASA) of the protein molecules were computed by the method of Richard and Lee using a probe size of 1.6 Å.[65] We employed CHARMM in computing SASA values of the proteins under given conditions.

Radial Distribution Function. The radial distribution function or the pair correlation function of B sites around A sites is given by

$$g_{\text{AB}}(r) = \frac{N_{\text{AB}}(r, \Delta r)}{\rho_{\text{B}} V_{\text{s}}(r, \Delta r)} \tag{10}$$

where, $N_{\text{AB}}(r, \Delta r)$ is the average number of B sites found in a shell, Δr thick, at a distance r from the A sites, V_{s} is the volume of this shell, and ρ_{B} is the average number density of B sites in the system. Different frames from the stable part of the trajectory are used to compute the radial distribution function of different atom pairs. Here we have used the radial distribution functions for comparison of the properties of water in the two cases.

The average number of water molecules H-bonded to the proteins was computed by CHARMM using the following criteria. A H-bond acceptor and donor pair is considered to form a hydrogen bond if the acceptor–hydrogen distance (d_{AH}) ≤ 2.6 Å and the donor–hydrogen–acceptor angle (θ_{DHA}) ≥ 135°.

RESULTS

We have used the mesophilic cold shock protein from *Bacillus subtilis* (*Bs*-CspB)[48] and the thermophilic cold shock protein from *Bacillus caldolyticus* (*Bc*-Csp),[49] as the natural mesophilic–thermophilic protein pair in the present investigation. The crystal structures (1csp, 1c9o) as well as the respective optimal growth temperatures ($T_{\text{opt}}^{\text{Meso}}$ = 35.0 °C and $T_{\text{opt}}^{\text{Thermo}}$ = 72.0 °C) of *Bs*-CspB and *Bc*-Csp are known.[50,51] *Bs*-CspB (67 residues) and *Bc*-Csp (66 residues) are small proteins without disulfide bonds. Each of the two structures contains a central beta-barrel motif as the main ordered structural unit. They differ in their amino acid sequence at 12 positions (Figure 1), all of which are on the protein surface. Thus, the two proteins of the pair are very similar in sizes and in amino acids sequences. It has been demonstrated experimentally that most of the different amino

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

Table 1. Summary of the Different Average Properties Obtained from the MD Simulations of the Mesophilic and Thermophilic Proteins at Different Temperatures

	Traj1 Thermo at T_{opt}^{Thermo}	Traj2 Meso at T_{opt}^{Meso}	Traj3 Thermo at T_{opt}^{Meso}	Traj4 Meso at T_{opt}^{Thermo}
average RMSD (Å) over last 4 ns (non-H atoms)	2.57 ± 0.23 Å	3.26 ± 0.18 Å	2.19 ± 0.13 Å	3.81 ± 0.15 Å
average self-energy (kcal/mol) over last 4 ns	−98.8 ± 66.3 kcal/mol	224.5 ± 63.6 kcal/mol	−138.6 ± 50.2 Å kcal/mol	266.6 ± 56.8 kcal/mol
average overall fluctuations over last 4 ns	1.48 Å	1.47 Å	1.10 Å	1.67 Å
average protein−solvent interaction energy (kcal/mol)	−1875 ± 134 kcal/mol	−2235 ± 154 kcal/mol	−1895.9 ± 93.1 kcal/mol	−1954.6 ± 110.5 kcal/mol
average SASA (Å^2) over last 4 ns (non-H atoms)	4620 ± 85 Å^2	4879 ± 76 Å^2	4510 ± 106 Å^2	4715 ± 90 Å^2

acids contribute to the observed enhanced stability as dictated by the melting temperatures T_m^{Meso} = 53.9 °C and T_m^{Thermo} = 76.9 °C.[66] The average properties of the two proteins obtained from the MD simulations at the respective optimal growth temperatures and at the cross optimal growth temperatures are summarized in Table 1.

Comparison of the Time Evolution of the 3D Structures. The time evolutions of the root mean squared deviation (RMSD) of the proteins from the respective initial structures in the MD simulations at the respective optimal growth temperatures are shown in Figure 2a. It is clearly indicated that the dynamical and structural features are similar for the thermophilic and mesophilic proteins of the pair at their respective optimal growth temperatures. The RMSD values over the last 4.0 ns appear to be reasonably stable and have average values 3.26 (±0.18) and 2.57 (±0.23) Å for the mesophilic and thermophilic variants respectively (Table 1). Thus, both of the proteins remained close to their respective initial structures and the thermophilic protein appears to remain more close to its initial crystal structure compared to that in the case of the mesophilic homologue.

Figure 2b compares the time evolutions of the RMSD values for the thermophilic protein and its mesophilic counterpart at their cross optimal growth temperatures. It is seen that the difference between the RMS deviations of the two proteins from their respective initial structures is quite large. The average RMSD of the trajectory of the mesophilic protein simulated at T_{opt}^{Thermo} is 3.81 ± 0.15 Å. Similarly, the average RMSD of the thermophilic protein simulated at T_{opt}^{Meso} is 2.19 ± 0.13 Å (Table 1). Thus, the structures are significantly different from the structures at the respective optimal growth temperatures.

For a direct comparison between the structures of the mesophilic and the thermophilic protein pair at their respective optimal growth temperatures, we computed the C-alpha atom based RMSD value between the structures averaged over the last 4 ns of the respective trajectories simulated at their respective optimal growth temperatures. The RMSD value between the trajectory averaged structures of the mesophilic and thermophilic proteins was found to be 1.61 Å and the same after refinement of the average structures by 1000 Steepest Descent energy minimization steps in vacuum was found to be 1.67 Å. Thus, at least the backbone structures in the two cases over the last 4 ns trajectories are found to be very similar. The average RMSD values in Table 1 indicate that the mesophilic protein at T_{opt}^{Thermo} has deviated significantly more with reference to its initial structure compared to that of the thermophilic homologue at T_{opt}^{Meso} and hence adopted a more floppy structure. Figure 3 compares the overall 3D structures of the two proteins.

Comparison of the Time Evolution of the Self-energies. Comparison of the time course of self-energies in the MD

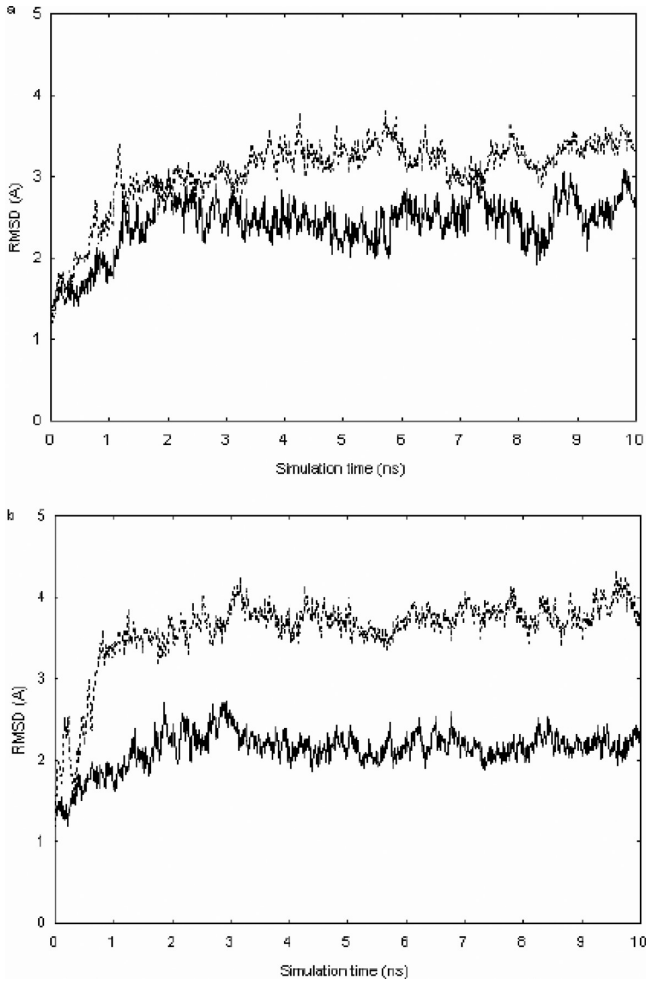


Figure 2. Comparison of the time evolutions of the RMSD values of the non-H atoms (a) of the mesophilic protein (dotted line) and the thermophilic counterpart (solid line) at their respective optimal growth temperatures T_{opt}^{Meso} (35 °C) and T_{opt}^{Thermo} (72 °C) and (b) of the mesophilic protein (dotted line) and the thermophilic counterpart (solid line) at their cross optimal growth temperatures T_{opt}^{Thermo} (72 °C) and T_{opt}^{Meso} (35 °C). Only frames at a gap of 4 ps have been used for visual clarity.

simulations of the thermophilic and mesophilic proteins at their respective optimal growth temperatures indicates that both the mesophilic and the thermophilic proteins remained stable throughout the period of simulation (Figure 4a). The average value of the self-energies over the last 4 ns of the respective trajectories for the mesophilic variant is 224.5 ± 63.6 kcal/mol while that for the thermophilic counterpart is −98.8 ± 66.3 kcal/mol (Table 1). This indicates that even under dynamical conditions, the thermophilic protein at T_{opt}^{Thermo} (72.0 °C),

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

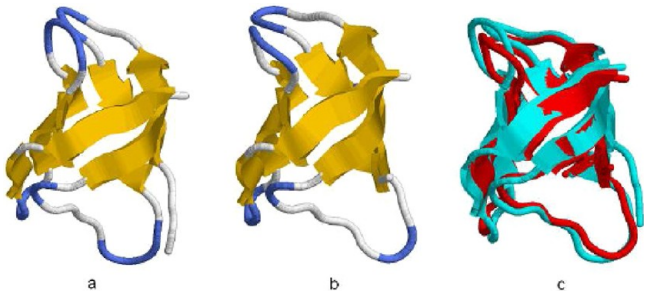


Figure 3. 3D structures in cartoon representation of (a) the mesophilic protein and (b) the thermophilic protein averaged over the last 4 ns of the respective trajectories. (c) The superposition of the averaged structures of the mesophilic (cyan) and the thermophilic (red) proteins.

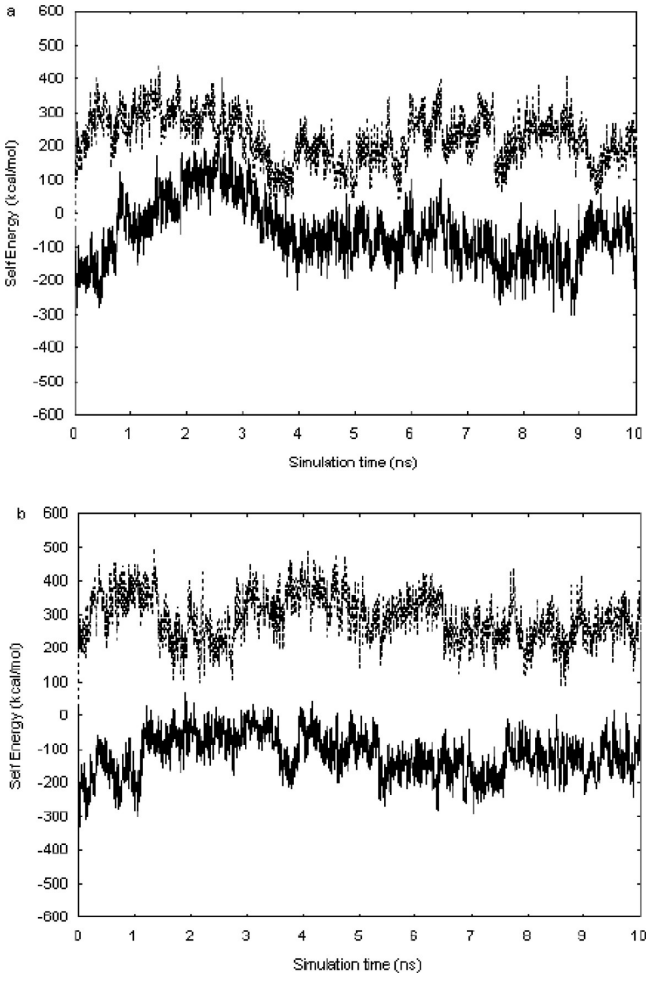


Figure 4. Time evolution of the self-energies of the mesophilic protein (dotted line) and the thermophilic counterpart (solid line) at (a) their respective optimal growth temperatures and (b) at their cross optimal growth temperatures. In both the cases, frames at a gap of 4 ps have been used for visual clarity.

remained more stable than the mesophilic protein at T_{opt}^{Meso} (35.0 °C). Thus, it is expected that the thermophilic variant may retain its structure even at a temperature higher than the one (72.0 °C) used here as its optimal growth temperature, and this is consistent with the experimental melting temperature 76.9 °C.[66] It may be mentioned here that the respective self-energies at t = 0 were −85.0 and −322.9 kcal/mol indicating much higher stability of the thermophilic protein. However, as mentioned earlier, the difference in the self-energies in such a case does not represent stability in an exact sense but is a reasonably good representative to compare the stability difference between two such proteins as explained in the Materials and Methods section.

For completeness, Figure 4b compares the time evolutions of the self-energies of the mesophilic and thermophilic proteins from their MD simulations at the cross optimal temperatures, and their average self-energies over the last 4 ns of the respective MD simulation trajectories are summarized in Table 1. The differences in the self-energies of each protein at the two temperatures demonstrate significant sensitivity of the self-energy values on temperature.

Comparison of Residuewise Average Fluctuations. We have considered the last 4 ns of each trajectory for computing the average root means squared fluctuations (RMSF) of the residues. RMSF values were calculated considering (i) all the atoms of a residue as well as (ii) the C-alpha atoms only. Figure 5a compares the residuewise RMS fluctuations of the

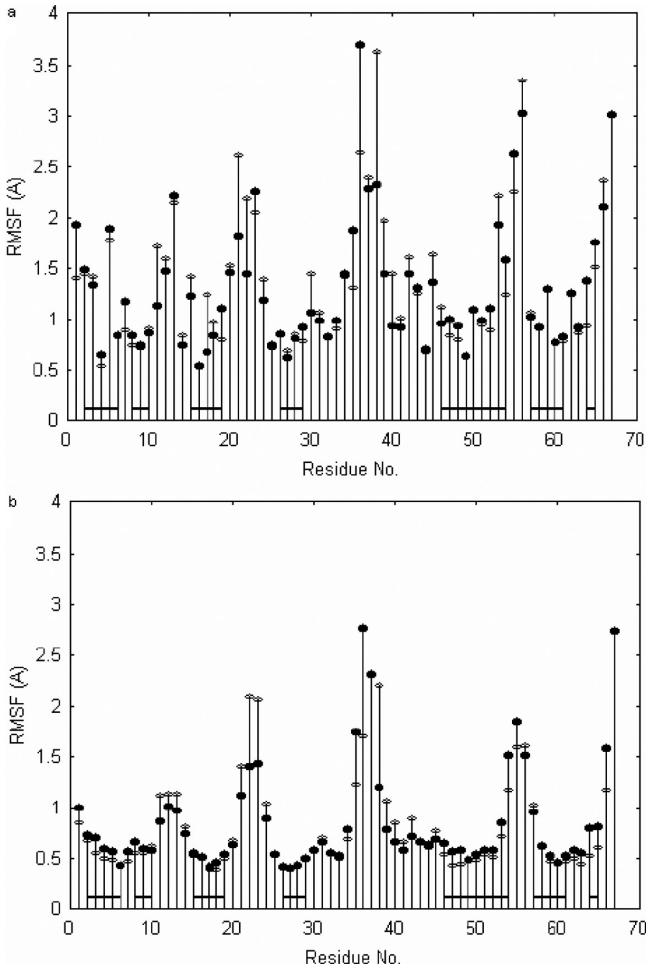


Figure 5. Comparison of the residuewise RMS fluctuations between the trajectories of the mesophilic protein (solid circles) and the thermophilic counterpart (open diamonds) at their respective optimal growth temperatures considering (a) all the atoms in a residue and (b) only the C-alpha atoms. The vertical lines are for guiding the eye to connect the residue number and the RMSF values easily. The short thick solid horizontal lines represent the residues of different regions that are involved in the beta-barrel structure.

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

thermophilic and mesophilic proteins from the two trajectories at their respective optimal growth temperatures, considering all the atoms in a residue. Primarily it is seen that the fluctuation patterns as well as the respective values are remarkably similar with a few exceptions. It is quite interesting to note that such high similarity is found in spite of the fact that the two proteins have identical amino acid sequences excepting only twelve residues and the temperatures at which the two MD simulations were carried out, are quite different. The computed values of $F_{overall}$ (5) for the thermophilic and the mesophilic proteins were found to be 1.48 and 1.47 Å respectively (Table 1) justifying the above-mentioned fact that these two proteins exhibit similar dynamics at their respective optimal growth temperatures in the time scale of tens of nanoseconds.

Excepting a few, the residuewise fluctuations for the thermophilic protein are generally found to be higher compared to that of the mesophilic counterpart. This is quite understandable because many residues are the same in the two proteins, and it is expected that they will show higher fluctuations at higher temperature T_{opt}^{Thermo} (72.0 °C) as happened for the thermophilic protein in comparison to the mesophilic case (T_{opt}^{Meso} = 35.0 °C). The residues at positions 1, 5, 11−13, 21−23, 35−39, 42, 45, 53−56, 65−66, and 67 (only for the mesophilic protein) are found to be higher (RMSF > 1.5 Å) compared to the other residues for both the proteins (Figure 5a). Examination of the 3D structures of the proteins indicates that these residues are in the loop regions exposed to solvent and hence fluctuate more. The same feature is also observed in the fluctuations of the C-alpha atoms and thus justifies our explanation (Figure 5b).

It is further interesting to note that many of the residues that are common between the two proteins exhibited similar fluctuations at the two significantly different temperatures. Figure 5a demonstrates that the residues at positions 4, 6−10, 16, 18−19, 25−29, 46−52, 57−58, 60−61, and 63 exhibit fluctuations smaller compared to that of the other residues in both the mesophilic and the thermophilic variants. This is consistent with the fact that all these residues are involved in a β barrel structure in each case.

Comparison of the fluctuations between the two proteins further indicates that the residues at the positions 11, 17, 21, 22, and 38−40 exhibit higher fluctuations in the thermophilic protein than that in the mesophilic protein. Among these residues only one residue (Ser-11) in the mesophilic protein was different (Asn-11) in the thermophilic one. When the thermophilic variant is simulated at T_{opt}^{Thermo} (72 °C), those common residues tend to execute higher fluctuation compared to those in the mesophilic one simulated at a lower temperature T_{opt}^{Meso} (35.0 °C). Among them which are on the surface of the proteins have the freedom for large fluctuation while the others which are in the interior of the protein have much lesser freedom due to packing. Thus, some of the residues common to both thermophilic and mesophilic variants exhibit higher fluctuations while some others show fluctuations similar to that in the mesophilic protein.

More interestingly, it is found that there are some residues at positions 1, 7, 19, 23, 35, 36, 54, 55, 64, and 65 that have lesser fluctuation in thermophilic variant compared to that of the mesophilic partner (Figure 5a). It is seen that some of these residues are the mutated residues or some residues close to a mutated residue. Such residues are, in general, interacting more strongly with the local part in the protein and thus are more stabilized in the thermophilc variant compared to that in the mesophilic counterpart (Table 2).

Table 2. Comparison of the Interaction Energies of the Residues Showing Lower Fluctuations Compared to the Rest of the Protein in the Thermophilic Protein than That in the Mesophilic Counterpart Simulated at Their Respective Optimal Growth Temperatures

residue no.	ΔRMSF = ($RMSF_T$ − $RMSF_M$)	E_m[b] (kcal/mol)	E_t[c] (kcal/mol)
1	−0.52	−69.7	−50.9
7	−0.27	−108.6	−126.4
19	−0.30	−54.2	−105.1
23[a]	−0.21	−18.2	−18.4
35	−0.56	−21.1	−24.2
36	−1.05	−21.1	−22.4
54	−0.34	−15.7	−18.9
55	−0.38	−15.5	−16.4
64[a]	−0.44	−27.4	−35.0
65	−0.24	−89.1	−71.8

[a]Signifies mutated residues in the thermophilic variant. [b]Average interaction energy of the residue with the rest of the protein in mesophilic variant. [c]Average interaction energy of the residue with the rest of the protein in thermophilic variant.

From Figure 5b, it is clearly seen that the fluctuations of the C-alpha atoms of the residues are very similar for the thermophilic and the homologous mesophilic protein excepting a few residues that are in the loop regions. This implies that the dynamics of the backbone parts are very similar in the two simulations.

In order to check how sensitive is the residuewise RMS fluctuation on the temperature, we have compared the residuewise average fluctuation values of the proteins under different temperatures. Figure 6a represents the comparison of the RMSF values of the mesophilic protein and the thermophilic counterpart at T_{opt}^{Meso}. The RMSF values are found to be considerably different. The residuewise RMSF values are found to be larger for the mesophilic protein at T_{opt}^{Meso} than that of the thermophilic protein at the same temperature (T_{opt}^{Meso}). The $F_{overall}$ value as defined before is 1.47 Å for mesophilic protein at T_{opt}^{Meso} while it is 1.10 Å for the thermophilic protein at the same temperature (T_{opt}^{Meso}). Thus, a significant difference in the overall RMSF values is observed between the mesophilic and thermophilic proteins simulated at the same temperature T_{opt}^{Meso} (35 °C). In a similar fashion, Figure 6b compares the residuewise RMSF values between the thermophilc protein at T_{opt}^{Thermo} (72 °C) and at T_{opt}^{Meso} (35 °C). The RMSF values are clearly seen to be significantly different. Here, the residuewise RMSF values for the thermophilic protein at T_{opt}^{Thermo} are found to be higher than that at T_{opt}^{Meso}. The $F_{overall}$ values for thermophilic protein simulated at T_{opt}^{Thermo} is 1.48 Å while it is 1.10 Å when simulated at T_{opt}^{Meso}. Thus, a significant difference in the overall RMSF values is found for the thermophilic protein simulated at the two different temperatures T_{opt}^{Thermo} (72 °C) and at T_{opt}^{Meso} (35 °C).

We have further compared the RMSF values of the residues for the mesophilic trajectory at T_{opt}^{Thermo} and the vice versa (Supporting Information Figure S1). It is seen that the average RMS fluctuation values of the residues of the two proteins are quite different. In general, significantly larger fluctuations are observed for the mesophilic trajectory compared to that of the

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

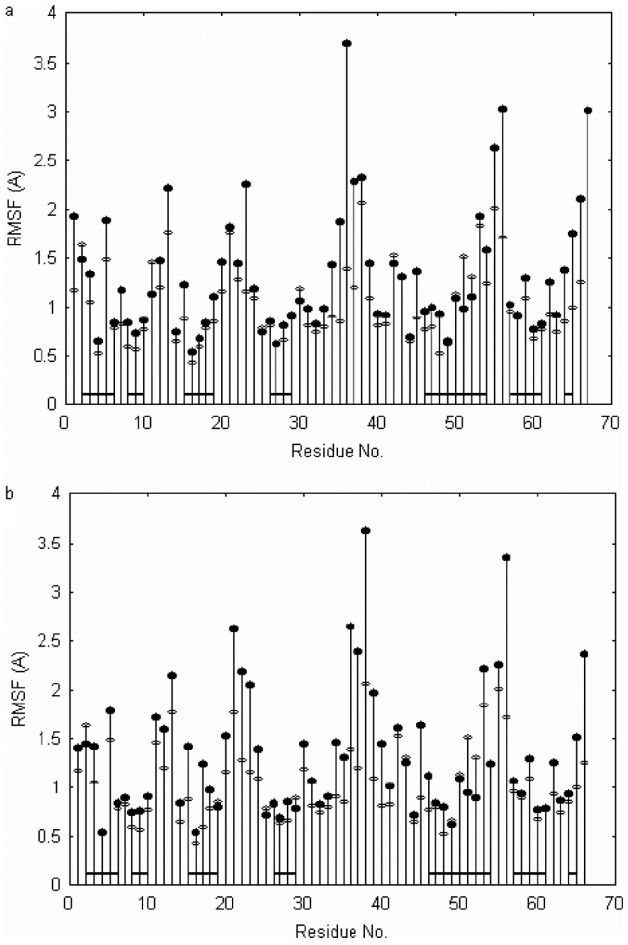


Figure 6. Comparison of the residuewise RMS fluctuations between the trajectories of (a) the mesophilic proteins at its optimal growth temperature T_{opt}^{Meso} (solid circles) and the thermophilic counterpart (open diamonds) at the same temperature T_{opt}^{Meso} considering all the atoms in a residue and (b) the thermophilic protein at its own optimal growth temperature T_{opt}^{Thermo} and at the optimal growth temperature T_{opt}^{Meso} of the mesophilic counterpart. The vertical lines are for guiding the eye to connect the residue number and the RMSF values easily. The short thick solid horizontal lines represent the residues of different regions that are involved in the beta-barrel structure.

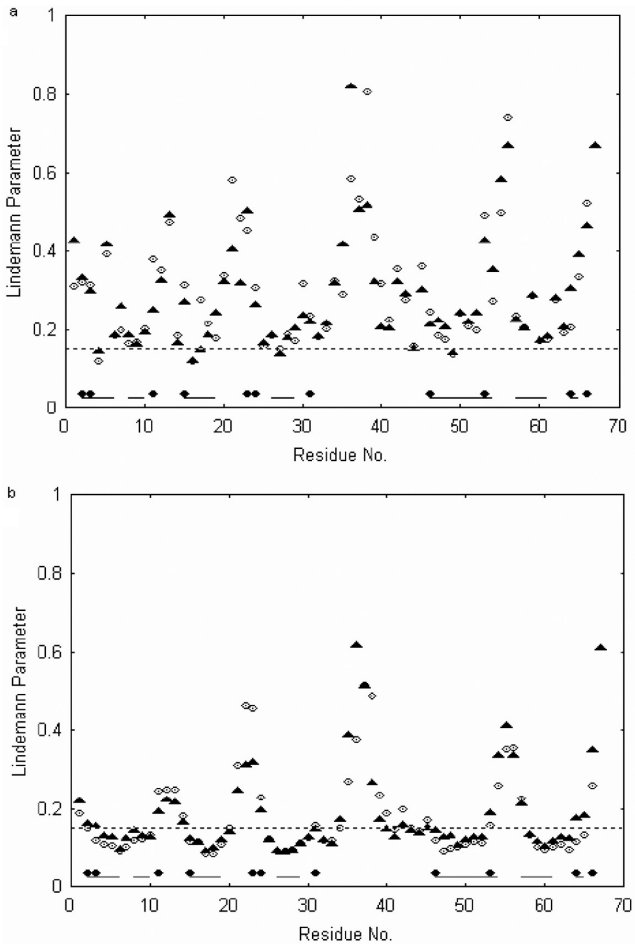


Figure 7. Comparison of the residuewise Lindemann parameter values between the trajectories of the mesophilic proteins (solid triangles) and the thermophilic counterpart (open circles) at their respective optimal growth temperatures considering (a) all the atoms in a residue and (b) only the C-alpha atoms. The dashed line represents the condition Δ_L^l = 0.15 Å. The short thick solid horizontal lines represent the residues of different regions that are involved in the beta-barrel structure. The solid circles represent the residues that are different in the thermophilic variant.

thermophilic counterpart. In contrast, the fluctuations obtained from the other trajectories, where the mesphilic protein and the thermophilic proteins were simulated at their respective optimal growth temperatures, are much more similar to each other (Figure 5a,b). All these observations support the view that at the optimal growth temperatures mesophilic and homologous thermophilic proteins exhibit similar fluctuations at least for the protein pair in the time scale (~10 ns) considered in this manuscript.

Solid-like and Liquid-like Characteristics of Residue Dynamics. It is interesting to assess the residuewise dynamics in terms of liquid-like and solid-like dynamics. The Lindemann criterion is very useful in characterizing such natures of the dynamics of a residue.[60−64] In order to characterize the dynamics of the residues of the thermophilic and mesophilic variants, we have computed the residuewise Lindemann parameter values (Δ_L^l) based on the relation 3 in the Materials and Methods section for both the mesophilic and thermophilic variants at respective optimal growth temperatures considering all atoms in a residue (Figure 7a). Application of the Lindemann criterion indicates that four residues at positions 4, 16, 27, and 49 in both the cases of the mesophilic and thermophilic proteins have Δ_L^l < 0.15 Å and thus behaved in a solid-like manner. All these residues are found to be involved in the central β barrel structure. The Lindemann criterion, applied to the fluctuation data for only the C-alpha atoms of the residues, indicates that 39 C-alpha atoms out of 67 (58%) exhibit solid-like dynamics for the mesophilic protein at 35 °C while the respective number is 37 out of 66 (56%) for the thermophilic simulation at 72 °C. Thus, for both the proteins at the respective optimal growth temperatures very similar fractions of the backbones exhibit solid-like dynamics. Figure 7b indicates that lesser number of the C-alpha atoms exhibit liquid-like dynamics and their Δ_L^l values are more close to solid-like character. Thus, these data suggest that at their optimal growth temperatures both the proteins behave as molten globules and both of them exhibit more or less similar liquid-like dynamics.

Comparison of the Δ_L^l values considering all atoms of the thermophilic protein at T_{opt}^{Meso}, with that of the mesophilic protein at T_{opt}^{Thermo} over the last 4 ns of the respective MD

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

simulations indicates that most of the residues of the mesophilic protein exhibit liquid-like behavior than the residues of the thermophilic protein in this case (Supporting Information Figure S2). Similar picture is also seen when the behaviors of the C-alpha atoms are compared (Supporting Information Figure S2). Thus, the mesophilic protein at T_{opt}^{Thermo} appears to be more floppy than the thermophilic protein at T_{opt}^{Meso} as observed by others in similar cases. However, the residues that are involved in the beta-barrel structure are found to exhibit solid-like dynamics in both the cases.

Figure 8 compares the number distribution of the Lindemann parameter (Δ_L^I) for the residues of the thermophilic

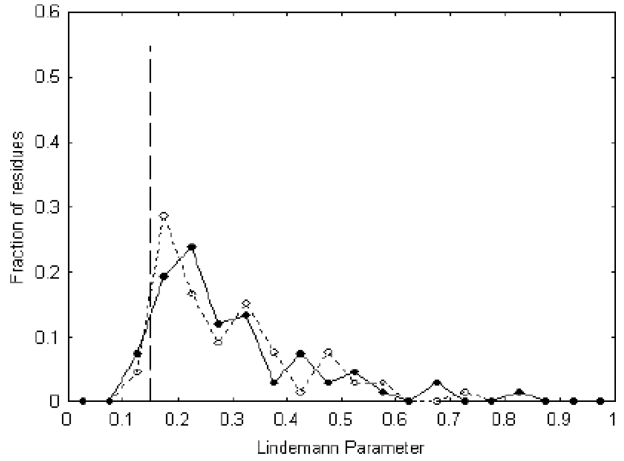


Figure 8. Comparison of the distribution of the computed Lindemann parameter (Δ_L^I) considering all the atoms in each residue for the mesophilic protein (solid circle and solid line) and the thermophilic counterpart (open circle and dotted line) at their respective optimal growth temperatures. The vertical dashed line represents the boundary at Δ_L^I = 0.15 Å.

and mesophilic proteins at respective optimal growth temperatures. It is seen that when all atoms are considered, about 33% and 27% of the residues have their Δ_L^I values within 0.20 Å for the thermophilic and mesophilic proteins respectively. Thus, for both the thermophilic and the mesophilic variants, a major number of residues have Δ_L^I values just outside the Lindemann criterion Δ_L^I < 0.15 Å for exhibiting solid-like dynamics. These residues are then expected to be liquid-like but very close to solid-like state. The Lindemann parameter values excepting a few are within the value of 0.45 Å implying that a significant number of residues are considerably liquid-like.

Comparison of Protein−Solvent Interaction and Solvent Behavior. The computed trajectory averaged value of the solvent accessible surface area over the last 4 ns of the trajectory of the thermophilic protein at T_{opt}^{Thermo} is found to be 4620 ± 85 Å^2 while the same for the mesophilic counterpart at T_{opt}^{Memo} is 4879 ± 76 Å^2 (Table 1). The initial values of the SASA for the thermophilic and mesophilic proteins are 4295 and 4630 Å^2, respectively. The difference in the initial SASA values for the thermophilic and the mesophilic proteins is mostly due to the fact that the mesophilic protein contains 67 residues while the thermophilic one consists of 66 residues. The average change in the SASA values for the thermophilic protein is 325 Å^2 where as, that for the mesophilic one is 249 Å^2. Thus, very similar changes in SASA with respect to the initial SASA values for the thermophilic and the mesophilic ones are observed.

The average protein−solvent interaction energies over the last 4 ns for the thermophilic and mesophilic proteins are −1875 ± 134 kcal/mol and −2235 ± 154 kcal/mol respectively. The mesophilic variant is thus seen to have more favorable interactions with the aqueous solvent compared to that in the case of the thermophilic protein. Similarly the average number of H-bonded water molecules is found to be 135 and 155 for the thermophilic and mesophilic variants, respectively. These are consistent with the expected higher dynamical fluctuations of water at T_{opt}^{Thermo} (72 °C which is higher than T_{opt}^{Meso}) for thermophilic protein MD simulation. Due to the higher fluctuations of the water molecules around the thermophilic protein, the average interactions between the protein and water became weaker compared to that of the mesophilic case.

In the cross temperature simulations, the interaction energies between the proteins are found to be −1895.9 ± 93.1 kcal/mol for the thermophilic protein and −1954.6 ± 110.5 kcal/mol for the mesophilic one. It is interesting to note that the solvent energy is more favorable in the case of the mesophilic protein than that of the thermophilic protein at the cross temperatures. This is because the mesophilic protein is floppy at the cross temperature and hence more residues are exposed to the solvent. Thus, more residues are interacting directly with solvent for the mesophilic protein than that in the case of the thermophilic protein at the cross temperatures.

Figure 9 compares the radial distribution function $g(r)$ plots for O−O, O−H, and H−H pairs for water molecules over the last 4 ns parts of the two trajectories at the respective optimal growth temperatures of the mesophilic and thermophilic proteins. Surprisingly, in each case, no significant differences between the trajectories of thermophilic and mesophilic proteins are observed. The peak heights for the thermophilic protein case are slightly smaller compared to those in the case of mesophilic one indicating that in the case of the thermophilic protein simulation there is some loss of ordered water molecules. Apart from this, the respective peak heights and peak positions in the cases of mesophilic and thermophilic trajectories appear quite similar. The peaks indicate that in both the cases there are significant amount of water structure and the differences in their distribution patterns are not much. This behavior is quite unexpected as the two trajectories were generated at two different temperatures and the temperature difference (37 °C) was significant, still the $g(r)$ plots do not show appreciable differences. This is most likely to be due to the fact that water molecules are strongly polar in nature and thus they interact strongly with each other. The temperature difference of 37 °C may not be large enough to introduce large structural and dynamical changes in the bulk water. Of course some small changes are clearly seen there in the plots (Figure 9). As a control experiment, we have performed MD simulations of pure Tip3P water box at the two temperatures 35 and 72 °C and found similar results (Supporting Information Figure S3) There are also other publications where water models other than Tip3P were used and reported to produce insignificant changes in various $g(r)$ plots for widely different temperatures.[67,68] If we combine these results with the results on protein−solvent interactions, it seems that water−water interaction is not significantly different but protein-water interaction is considerably different in the two cases.

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

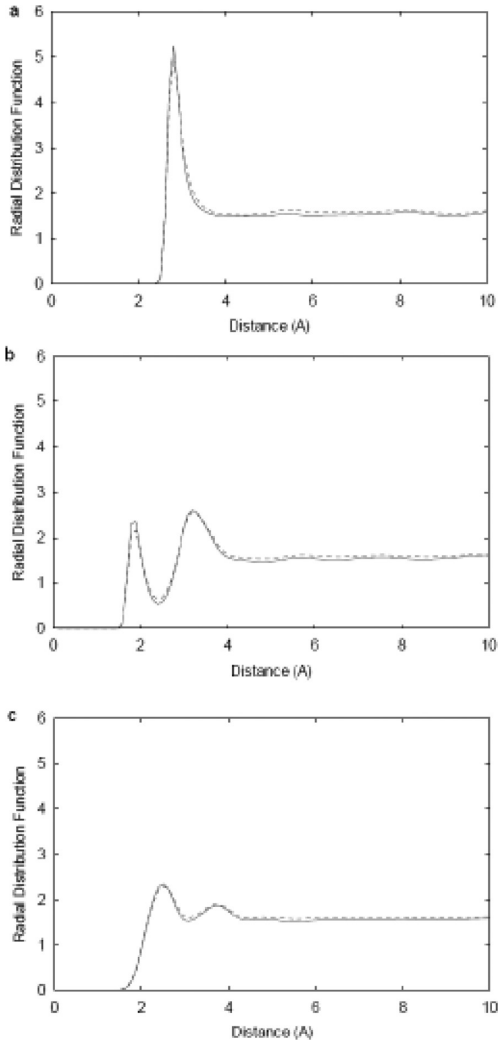


Figure 9. Comparison of the radial distribution functions of water obtained from the MD simulations of the mesophilic protein at T_{opt}^{Meso} (solid line) and the thermophilic counterpart at T_{opt}^{Thermo} (dotted line): (a) O−O pair, (b) O−H pair, and (c) H−H pair.

DISCUSSIONS

MD simulation studies on a naturally occurring homologous pair of mesophilic−thermophilic proteins have been carried out at the respective optimal growth temperatures and also at the cross temperatures and the trajectories have been analyzed. There are experimental evidence that in general thermophilic proteins are mostly inactive at T_{opt}^{Meso}.[69] This is consistent with the idea that the activity of a protein is not only determined by its 3D structure but also by its dynamics. In addition to emphasizing the above fact, present studies also reveal that the dynamical and structural features are remarkably similar for the partners of a homologous thermophilic and mesophilic protein pair at their respective optimal growth temperatures.

It may further be noted that the results discussed here are based on the MD simulations of 10 ns time scale. The dynamics at much larger time scale can be quite different. During a biologically relevant long time scale dynamics, the structure of the protein is changed significantly due to the segmental motions of the protein and as a result the immediate environment of different parts of the protein can be changed considerably. As a consequence, during such a large scale dynamics, short scale dynamics (tens of nanoseconds or so) can be significantly different as reported in some publications.[70,71] However, our results are confined within the 10 ns time scale only.

We may point out here that when the mesophilic protein is simulated at T_{opt}^{Thermo} (72 °C), the protein was expected to be denatured. However, as the simulation time scale is too short (only 10 ns) the protein is found to be floppy only. This is supported by the fact that the average interaction energy between the protein and solvent is much stronger in this case. For exhibiting significant denaturation one need to perform a very long simulation.

It is found here that the common residues on or close to the surface of the thermophilic protein at T_{opt}^{Thermo} are fluctuating more compared to that of its mesophilic homologue at T_{opt}^{Meso}. Thus, the chemical environment at the active site of thermophilic protein is expected to be in a more dynamical condition and hence will represent more average local properties than that of the mesophilic counterpart. As a result, the activity of a thermophilic protein at its high T_{opt}^{Thermo} is likely to be less compared to that of the mesophilic counterpart at T_{opt}^{Meso} because, in general, the active site residues remain the same in both the mesophilic and thermophilic homologues. On the other hand, the elevated temperature may directly improve the effective activity for specific chemical reactions. Higher temperature may also help in reaching the active conformation faster in thermophilic protein by facilitating the rate limiting segmental motions and thereby increasing the enzymatic efficiency. In that case, the end result will be determined by the net effects of these opposite contributions. Due to the unavailability of adequate experimental data on the activities of mesophilic−thermophilic protein pair, we could not check the validity of this possibility. However, there is a paper that reports the decrease in activity of the thermophilic enzyme compared to that of the mesophilic counterpart for a different protein.[72] Thus, it remains an interesting issue that needs thorough experimental and theoretical investigations in the future.

CONCLUSIONS

The MD simulation trajectories at the respective optimal growth temperatures indicate that the dynamical and structural features are remarkably similar for the thermophilic and the homologous mesophilic protein pair at their respective optimal growth temperatures.

Residuewise fluctuations indicate that for many residues the fluctuations are very similar in mesophilic and thermophilic variants. The residues which are in the loops are fluctuating more compared to the other residues. Residues showing lower fluctuations are in general parts of the central beta-barrel structure. Some mutated residues as well as some conserved residues close to a mutated residue have lesser fluctuations in the thermophilic protein than the corresponding residues in the mesophilic counterpart due to additional local stability caused by interactions between the mutated residues and some of the local residues in the thermophilic protein.

Comparison of Lindemann parameters indicates that both the proteins at respective optimal growth temperatures are in molten globule sates and the thermophilic protein is slightly more liquid-like compared to its mesophilic counterpart. Only a few residues of the thermophilic and mesophilic proteins behave in solid-like manner.

The radial distribution patterns of water molecules are surprisingly similar in the MD simulations of the thermophilic and mesophilic proteins. Slightly lower number of bound water

dx.doi.org/10.1021/ci300474h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

molecules is found in the thermophilic case compared to that of the mesophilic one.

It seems that the water−water interactions did not change significantly but the water−protein interaction changed noticeably.

■ ASSOCIATED CONTENT

Supporting Information

Comparison of the residuewise average RMSF values of the thermophilic protein at T_{opt}^{Meso} with that of the mesophilic protein at T_{opt}^{Thermo} is shown considering all atoms and separately considering only the C-alpha atoms in Figure S1. Similarly, comparison of the residuewise Lindemann parameters values between the trajectories of the mesophilic protein at T_{opt}^{Thermo} and the thermophilic counterpart at T_{opt}^{Meso} is presented in Figure S2. Figure S3 represents the comparisons of O−O, O−H, and H−H pair distributions in a pure water box at the two temperatures T_{opt}^{Meso} and T_{opt}^{Thermo}. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*Fax: +91 33 23574113. Tel.: +91 33 23572744/45. E-mail: srikanta@chembiotek.com, srikantasen@ymail.com.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1−43.

(2) de Champdoré, M.; Staiano, M.; Rossi, M.; D'Auria, S. Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest. *J. R. Soc. Interface* **2007**, *4*, 183−191.

(3) Rees, D. C.; Adams, M. W. W. Hyperthermophiles: Taking the heat and loving it. *Structure* **1995**, *3*, 251−254.

(4) D'Amico, S.; Marx, J. C.; Gerday, C.; Feller, G. Activity−stability relationships in extremophilic enzymes. *J. Biol. Chem.* **2003**, *278*, 7891−7896.

(5) Russell, R. J. M.; Ferguson, J. M. C.; Hough, D. W.; Danson, M, J.; Taylor, G. L. The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. *Biochemistry* **1997**, *36*, 9983−9994.

(6) Criswell, A. R.; Bae, E.; Stec, B.; Konisky, J.; Phillips, G. N., Jr. Structures of thermophilic and mesophilic adenylate kinases from the genus Methanococcus. *J. Mol. Biol.* **2003**, *330*, 1087−1099.

(7) Chien, A.; Edgar, D. B.; Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *J. Bacteriol.* **1976**, *127*, 1550−1557.

(8) Bruins, M. E.; Janssen, A. E.; Boom, R. M. Thermozymes and their applications: A review of recent literature and patents. *Appl. Biochem. Biotechnol.* **2001**, *90*, 155−186.

(9) Turner, P.; Mamo, G.; Karlsson, E. N. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb. Cell. Fact.* **2007**, *6*, 9−31.

(10) Lehmann, M.; Loch, C.; Middendorf, A.; Studer, D.; Lassen, S.; Pasamontes, L.; van Loon, A. P. G. M.; Wyss, M. The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **2002**, *15*, 403−411.

(11) Dantas, G.; Kuhlman, B.; Callender, D.; Wong, M.; Baker, D. A Large Scale Test of Computational Protein Design: Folding and Stability of Nine Completely Redesigned Globular Proteins. *J. Mol. Biol.* **2003**, *332*, 449−460.

(12) Eijsinka, V. G. H.; Gaseidnesa, S.; Borchert, T. V.; van den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **2005**, *22*, 21−30.

(13) Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **2005**, *308*, 857−860.

(14) Zollars, E. S.; Marshall, S. A.; Mayo, S. L. Simple electrostatic model improves designed protein sequences. *Protein Sci.* **2006**, *15*, 2014−2018.

(15) Montanucci, L.; Fariselli, P.; Martelli, P. L.; Casadio, R. Predicting protein thermostability changes from sequence upon multiple mutations. *Bioinformatics* **2008**, *24*, i190−i195.

(16) Bannen, R. M.; Suresh, V.; Phillips, G. N., Jr.; Wright, S. J.; Mitchell, J. C. Optimal design of thermally stable proteins. *Bioinformatics* **2008**, *24*, 2339−2343.

(17) Basu, S.; Sen, S. Turning a mesophilic Protein into a Thermophilic one: A Computational Approach based on 3D structural features. *J. Chem. Inf. Model.* **2009**, *49*, 1741−1750.

(18) Basu, S.; Sen, S. Designing thermophilic proteins: A structure-based computational approach. In *Thermostable Proteins: Structural Stability and Design*; Sen, S., Nilsson, L., Eds.; Francis & Taylor: CRC Press, USA, 2011; Chapter-6, pp 117−153.

(19) Vogt, G.; Woell, S.; Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **1997**, *269*, 631−643.

(20) Karshikoff, A.; Ladenstein, R. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **1998**, *11*, 867−872.

(21) Grimsley, G. R.; Shaw, K. L.; Fee, L. R.; Alston, R. W.; Huyghues-Despointes, B. M.; Scholtz, J. M.; Pace, C. N. Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **1999**, *8*, 1843−1849.

(22) Jaenicke, R. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2926−2964.

(23) Szilagyi, A.; Zavodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comparative survey. *Structure* **2000**, *8*, 493−504.

(24) Kannan, N.; Vishveshwara, S. Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13*, 753−761.

(25) Perl, D.; Schmid, F. X. Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **2001**, *313*, 343−357.

(26) Dominy, B. N.; Minoux, H.; Brooks, C. L., 3rd An electrostatic basis for the stability of thermophilic proteins. *Proteins* **2004**, *5*, 128−141.

(27) Thomas, A. S.; Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J. Am. Chem. Soc.* **2004**, *126*, 2208−2214.

(28) Razvi, A.; Scholtz, J. M. Lessons in stability from thermophilic proteins. *Protein Sci.* **2006**, *15*, 1569−1578.

(29) Gribenko, A. V; Makhatadze, G. I. Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J. Mol. Biol.* **2007**, *366*, 842−856.

(30) Kumar, S.; Tsai, C. J.; Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **2000**, *13*, 179−191.

(31) Chakravarty, S.; Varadarajan, R. Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study. *Biochemistry* **2002**, *41*, 8152−8161.

(32) Gianese, G.; Bossa, F.; Pascarella, S. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins* **2002**, *47*, 236−249.

(33) Sadeghi, M.; Naderi-Manesh, H.; Zarrabi, M.; Ranjbar, B. Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **2006**, *119*, 256−270.

(34) Greaves, R. B.; Warwicker, J. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. *BMC Struct. Biol.* **2007**, *7*, 18−40.

(35) Kumar, S.; Ma, B.; Tsai, C. J.; Nussinov, R. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* **2000**, *38*, 368−383.
(36) Makhatadze, G. I.; Loladze, V. V.; Ermolenko, D. N; Chen, X. F.; Thomas, S. T. Contribution of surface salt bridges to protein stability: Guidelines for protein engineering. *J. Mol. Biol.* **2003**, *327*, 1135−1148.
(37) Grottesi, A.; Ceruso, M. A.; Colosimo, A.; Di Nola, A. Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. *Proteins* **2002**, *46*, 287−294.
(38) Bae, E.; Phillips, G. N., Jr. Identifying and Engineering Ion Pairs in Adenylate Kinases Insights from molecular dynamics simulations of thermophilic and mesophilic homologues. *J. Biol. Chem.* **2005**, *280*, 30943−30948.
(39) Melchionna, S.; Raffaele Sinibaldi, R.; Briganti, G. Explanation of the Stability of Thermophilic Proteins Based on Unique Micromorphology. *Biophys. J.* **2006**, *90*, 4204−4212.
(40) Sterpone, F.; Bertonati, C.; Briganti, G.; Melchionna, S. Key role of proximal water on thermostable proteins. *J. Phys. Chem. B.* **2009**, *113*, 131−137.
(41) Merkley, E. D.; William, W.; Parson, W. W.; Daggett, V. Temperature dependence of the flexibility of thermophilic and mesophilic flavoenzymes of the nitroreductase fold. *Prot. Eng., Des. Selec.* **2010**, *23*, 327−336.
(42) Tiberti, M.; Papaleo, E. Dynamic properties of extremophilic subtilisin-like serine-proteases. *J. Struct. Biol.* **2011**, *174*, 69−82.
(43) Marcos, E.; Jimenez, A.; Crehuet, R. Dynamic Fingerprints of Protein Thermostability Revealed by Long Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1129−1142.
(44) Lee, K. J. Molecular Dynamics Simulations of a Hyperthermophilic and a Mesophilic Protein L30e. *J. Chem. Inf. Model.* **2012**, *52*, 7−15.
(45) Sterpone, F.; Melchionna, S. Thermophilic proteins: insight and perspective from *in silico* experiments. *Chem. Soc. Rev.* **2012**, *41*, 1665−1676.
(46) Merkley, E. D.; W.Parson, W. W.; Valerie Daggett, V. Temperature dependence of the flexibility of thermophilic and mesophilic flavoenzymes of the nitroreductase fold. *Prot. Eng. Des. Selec.* **2010**, *23*, 327−336.
(47) Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys.: Condens. Matter* **2010**, *22*, 323101.
(48) Schindelin, H.; Marahiel, M. A.; Heinemann, U. Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. *Nature* **1993**, *364*, 164−168.
(49) Müller, U.; Perl, D.; Schmid, F. X; Heinemann, U. Thermal stability and atomic-resolution crystal structure of the Bacillus caldolyticus cold shock protein. *J. Mol. Biol.* **2000**, *297*, 975−988.
(50) Heinen, W.; Lauwers, A. M. Growth of bacteria at 100°C and beyond. *Arch. Microbiol.* **1981**, *129*, 127−128.
(51) Kumar, S.; Tsai, C. J.; Nussinov, R. Thermodynamic Differences among Homologous Thermophilic and Mesophilic Proteins. *Biochemistry* **2001**, *40*, 14152−14165.
(52) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Kleon, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(53) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(54) York, D. M.; Darden, T. A.; Pedersen, L. G. The effect of longrange electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(55) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667−683.
(56) Norberg, Y.; Nilsson, L. On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophys. J.* **2000**, *79*, 1537−1553.
(57) Cheatham, T. E.; Kollman, P. E. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.
(58) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(59) MacKerell, A. D., Jr.; Bashford, D.; Dunbrack, M. R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **1998**, *102*, 3586−3616.
(60) Zhou, Y.; Vitkup, D.; Karplus, M. Native Proteins are Surface-molten Solids: Application of the Lindemann Criterion for the Solid versus Liquid State. *J. Mol. Biol.* **1999**, *285*, 1371−1375.
(61) Bilgram, J. H. Dynamics at solid liquid transition - experiments at the freezing-point. *Phys. Rep.* **1987**, *153*, 1−89.
(62) LoÈwen, H. Melting, freezing and colloidal suspensions. *Phys. Rep.* **1994**, *237*, 249−324.
(63) Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **1995**, *267*, 1935−1939.
(64) Stillinger, F. H.; Stillinger, D. K. Computational study of transition dynamics in 55-atom clusters. *J. Chem. Phys.* **1990**, *93*, 6013−6024.
(65) Lee, B. K.; Richards, F. M. The interpretation of protein structures. Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(66) Perl, D.; Mueller, U.; Heinemann, U.; Schmid, F. X. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* **2000**, *7*, 380−383.
(67) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, non-polarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910−8920.
(68) Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000**, *258*, 121−137.
(69) Hecht, K.; Wrba, A.; Jaenicke, R. Catalytic properties of thermophilic lactate dehydrogenase and halophilic malate dehydrogenase at high temperature and low water activity. *Eur. J. Biochem.* **1989**, *183*, 69−74.
(70) Butterwick, J. A.; Loria, J. P.; Astrof, N. S.; Kroenke, C. D.; Cole, R.; Rance, M.; Palmer, A. G., III. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J. Mol. Biol.* **2004**, *339*, 855−871.
(71) Meinhold, L.; Clement, D.; Tehei, M.; Daniel, R.; Finney, J. L.; Smith, J. C. Protein Dynamics and Stability: The Distribution of Atomic Fluctuations in Thermophilic and Mesophilic Dihydrofolate Reductase Derived Using Elastic Incoherent Neutron Scattering. *Biophys. J.* **2008**, *94*, 4812−4818.
(72) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Biol.* **2004**, *11*, 945−949.