JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

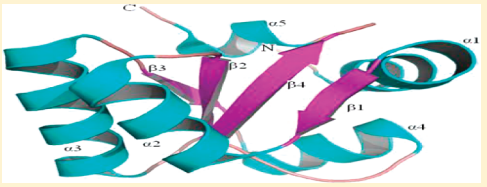
pubs.acs.org/jcim

Molecular Dynamics Simulations of a Hyperthermophilic and a Mesophilic Protein L30e

Kuei-Jen Lee*

Department of Bioinformatics, Asia University, Taichung, Taiwan 413, Republic of China

ABSTRACT: Molecular dynamics (MD) simulations were used to study the hyperthermophilic ribosomal protein L30e from archaeon *Thermococcus celer* at 300 and 350 K, and its mesophilic homologue, yeast L30e, at 300 K in explicit solvent for a period of 5.0 ns. Three trajectories obtained from the MD simulations were stable throughout the simulation period, such as total potential energy, radius of gyration, root-mean-square deviation, and secondary structures assignment. At 300 K, *T. celer* L30e is less flexible than its mesophilic homologue, and this difference becomes more pronounced at 350 K. Salt bridge networks, one triad and one hexad, are present at the surface of *T. celer* L30e. The ion pairs forming these salt bridges maintain close contact at a higher temperature, suggesting that these networks contribute to the protein's hyperthermal stability. By contrast, we found no such networks to be present in yeast L30e. For charged residue I in *T. celer* L30e, the $\Delta\Delta G_{solv}{}^{I}$ value and its corresponding $\Delta E_{Coul}{}^{I}$ value possess opposite signs. This indicates that for *T. celer* L30e, a change in the solvation free energy of a charged residue due to increasing temperature is compensated by a change in the residude's Coulombic interaction energy with the rest of the protein.



INTRODUCTION

Organisms with an optimal growth temperature between 80 and 110 °C are termed hyperthermophiles, while thermophilic organisms grow optimally between 50 and 80 °C.[1] Optimal growth temperatures for mesophilic organisms range from 20 to 50 °C.[1] Proteins isolated from thermophilic or hyperthermophilic organisms remain stable and active at extreme temperatures, while their mesophilic homologues become denatured.[2,3] Thermophilic or hyperthermophilic proteins and their mesophilic homologues usually share a high degree of similarity in their sequences and three-dimensional (3D) structures.[4,5] Proteins that are stable at high temperatures have attracted much interest because they have potential industrial applications;[6,7] thus, it is important to understand how these thermophilic proteins remain stable at elevated temperatures. Such an understanding may help to elucidate critical principles of protein engineering and inform the design of thermostable proteins for industrial applications. Comparative studies[4,8−17] have revealed several prominent feactures often exhibited by thermophilic proteins compared to their mesophilic homologues, including increased hydrogen bonds,[9,10,14] salt bridges,[9−15] fewer thermolabile residues,[4] decreased number and volume of internal cavities,[12] increased hydrophobicity[8,13,16] in protein cores, and increased interface stabilization between oligomeric protein subunits.[17] These studies are based on 3D protein structures available in the Protein Data Bank (PDB) or that have been predicted using homology modeling[18] of thermophilic/mesophilic pairs.[9,13] The structures of proteins, therefore, play an important role in governing their thermostability. It seems that there are no general rules to determine thermostability, and it is apparent that different proteins attain their thermostabity due to different combinations of the above factors.

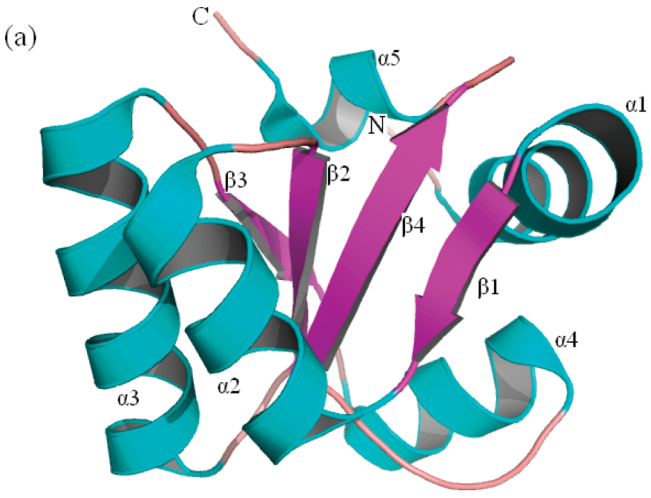
A whole genome analysis of various hyperthermophiles[19] has revealed a large proportion of charged residues to be present in hyperthermophilic proteins. According to comparative studies, many thermophilic proteins have an increased number of salt bridges[9−15] compared to the analogous mesophilic proteins. Furthermore, there are a large number of salt bridges present on the surface of thermophilic proteins, and these salt bridges may participate in a complex interplay[20,21] at intraprotein, interdomain, and intersubunit interfaces of oligomeric proteins. Thus, temperature dependency of molecular structures and their dynamics may be factors in charged residue interaction. Molecular dynamics simulation at varying temperatures[22,23] is a useful tool to investigate the dynamic properties of charged residue interactions at the atomic level. Investigations of the structure and property relationships[24] of proteins at the atomic level, which sometimes proves difficult to achieve in experimental studies, are important because thermostable properties are influenced by protein molecular structures.

This study primarily focuses on the hyperthermophilic ribosomal protein L30e from archaeon *Thermococcus celer*[25] and its mesophilic homologue from the yeast *Saccharomyces cerevisiae*.[26] In yeast, L30e, formerly known as L32,[27,28] is a ubiquitous component of a large subunit in the eukaryotic and archaeal ribosomes, though it has no homologue in prokarya. In yeast, L30e binds to a purine-rich internal loop in its pre-mRNA and mRNA, to autoregulate splicing and translation functions, respectively.[29−31] *T. cele* L30e is a single domain protein composed of 100 residues,[25] while yeast L30e consists of 104 residues.[26] The most significant difference between the two proteins is in their thermostabilities. The melting temperature of *T. celer* L30e is ∼94 °C,[25] while that of yeast

Received: April 27, 2011
Published: December 15, 2011

dx.doi.org/10.1021/ci200184y | *J. Chem. Inf. Model.* 2012, 52, 7−15

is 46 °C.[26] *T. celer* L30e exhibits reversible thermal denaturation,[25] whereas yeast L30e unfolds irreversibly at temperatures greater than 45 °C.[25] The structure of *T. celer* L30e was determined both by nuclear magnetic resonance (NMR) spectroscopy[25] and by X-ray crystallography,[32] whereas the structure of yeast L30e was only investigated by NMR spectroscopy[26] in both free and pre-mRNA bound forms. Both proteins present an $\alpha/\beta/\alpha$ sandwich pattern (Figure 1a and b). The topology of



(c)

```
T.celer   1 ------VDFAFELRKAQDTGKIVMGARKSIQYAKMGGAKLIIVARNARPDIKEDIEYYAR---54
Yeast     2 APVKSQESINQKLALVIKSGKYTLGYKSTVKSLRQGKSKLIIAANTPVLRKSELEYYAM---61

T.celer  55 LSGIPVYEFEGTSVELGTLLGRPHTVSALAVVDPGESRILALGGKE-----100
Yeast    62 LSKTKVYYFQGGNNELGTAVGKLFRVGVVSILEAGDSDILTTLA-------105
```

Figure 1. (a) Structure of *T. celer* L30e from NMR structure (1GO1).[25] (b) Structure of yeast L30e from NMR structure (1CN7).[26] Figures in both (a) and (b) are drawn using the program PyMOL (www.pymol.org). The secondary structure elements in both (a) and (b) are also labeled. N and C in both (a) and (b) indicate the amino and carboxy termini, respectively. (c) The amino acid sequences of *T. celer* L30e and yeast L30e. The sequences were aligned using the CLUSTAL W program (www.ebi.ac.uk/clustalw). The residues in *T. celer* L30e and yeast L30e are numbered according to those reported in NMR structures.

T. celer L30e contains four α-helices and one small 3_{10} helix, as indicated by α5 (Figure 1a), and a mixed β-sheet comprising of four β-strands (Figure 1a).[25] Yeast L30e contains four α-helices and four β-strands (Figure1b).[26] Although the aligned amino acid sequences of *T. celer* L30e and yeast L30e show that both proteins share about 30% sequence identity (Figure 1c), the structures of these two proteins are very similar, and the root-mean-square deviation (rmsd) of backbone atoms between these two proteins is about 2.0 Å.[25]

Previous investigators used MD simulations to investigate unfolding behavior of homologous proteins via a series of temperatures[33] to address the problem of activity and thermostability of mesophilic subtilisin E and its thermophilic mutant,[34] to investigate the internal flexibility of thermophilic/mesophilic pairs,[35,36] and to convert a mesophilic protein to a thermophilic protein.[37] In this study, MD simulations of two L30e proteins from different sources were carried out to investigate internal dynamic fluctuations and the thermal stability of the proteins native fold at two different temperatures (300 and 350 K) for the hyperthermophile *T. celer* L30e and at 300 K for its mesophilic homologue yeast L30e. The dynamic properties of each protein were compared to probe the key factors governing the hyperthermal stability of *T. celer* L30e at 300 K. The electrostatic contributions to the thermal stability of the hyperthermophilic L30e via salt bridges and networks of ion pairs were examined, and the electrostatic solvation free energy and Coulombic interaction energy for individual charged residues was also investigated.

MATERIALS AND METHODS

Molecular Models. The NMR structures for both *T. celer* L30e and yeast L30e proteins were obtained from the Protein Data Bank (PDB). The entry codes 1GO1[25] and 1CN7[26] (RNA-free structure) for *T. celer* L30e and for yeast L30e, respectively, were used as the initial models for our MD simulations. The protonation states of ionizable residues at pH 7.0 were selected based on pK_a values of the isolated side-chain groups, while N and C termini were considered to be ionized. *T. celer* L30e has 13 acidic residues and 14 basic residues, resulting in a net total charge of +1 e. Yeast L30e has 7 acidic residues and 15 basic residues, having a net total charge of +8 e. In addition, metal cofactors are assumed to be absent from these proteins, and they are of a small size (~11 kD),[25,26] thereby making *T. celer* L30e and yeast L30e ideal proteins for comparative studies by MD simulation.

MD Simulation. All MD simulations were carried out using the CHARMM[38] program with the all-atom parameter set PARAM22 for proteins.[39] Hydrogen atoms were added using HBUILD facility in CHARMM. The resulting L30e protein structures were minimized in vacuo without constraints, using 2000 steps of steepest decent (SD) to eliminate bad atomic contacts and then further minimized using 2000 steps of the Adopted Basis Newton−Raphson (ABNR) algorithm in CHARMM. The overall rmsd of the minimized *T. celer* L30e and yeast L30e from their starting structures was 0.361 Å and 0.364 Å, respectively. The energy-minimized proteins were solvated in a pre-equilibrated cubic box of TIP3 water molecules[40] with dimensions 52.7 × 52.7 × 52.7 Å. We randomly replaced water molecules close to Lys or Arg residues in the proteins with chloride ions to neutralized the +1 charge of 1GO1 and the +8 charge of 1CN7. Thus, the solvated systems resulted in a total of 4291 TIP3P water molecules around 1GO1 and 4283 water molecules around 1CN7. The resulting solvated systems were subjected to SD minimization for 2000 steps under harmonic restraints, with 50 kcal/mol/Å^2 for the

protein heavy atoms; ABNR minimization followed, with 20 kcal/mol/Å^2 constraints for 3000 steps. This resolved poor contacts between solvent and solute (proteins and ions) without disrupting the overall protein conformations. The protein's heavy atoms were then fixed with 5 kcal/mol/Å^2 constraints, and both solvent and ions were SD minimized, using 2000 steps followed by a further 3000 steps of ABNR minimization for each system, without solute constraints. Both systems were heated to 300 or 350 K target temperatures, starting from 50 K. The temperature was increased in increments of 1.0 K at every 100 integration steps, using a time step of 1 fs until 300 or 350 K was achieved. Finally, the systems were equilibrated for 25 ps at 300 or 350 K with thermal fluctuations within 5 K and with velocity rescaling every 0.05 ps with a 1 fs time step. Following equilibration, the simulations were performed on the NPT ensemble set to 1 atm pressure, using the CPT algorithm[41] and employing the Leap-Frog integrator method. All hydrogen atom bond lengths were constrained using SHAKE[42] with a tolerance of 1.0×10^{-6} Å. Periodic boundary conditions were applied to the systems to reduce edge effects. Long-range electrostatic interactions were computed using the Particle−Mesh−Ewald (PME) method,[43] in which a β-spline order of 6, a 60 × 60 × 60 grid for fast Fourier transforms (FFTs), and a real space Gaussian width, κ, of 0.32 Å^{-1} were used; a unit dielectric constant was used in the calculation of electrostatic interactions. The list of nonbonding interactions was truncated at 14 Å. The nonbonding van der Waals and real-space electrostatic interaction functions were switched smoothly on and off from 8 to 10 Å, respectively.[44] The MD simulations were carried out for 300 or 350 K over 5.0 ns and using a 2 fs time step. Coordinates were saved every 1.0 ps.

Trajectory Analysis. The MD simulations provided the total potential energy[38,39] of the simulated systems, E_{total}. The most important contributions to the total potential energy are due to bonding and nonbonding energies, and these are represented as eqs 1 and 2

$$E_{bond} = E_b + E_\theta + E_{UB} + E_\phi + E_\omega \quad (1)$$

and

$$E_{nonbond} = E_{vdW} + E_{Coul} \quad (2)$$

where E_b, E_θ, E_{UB}, E_φ are the bond stretching, bond angle bending, Urey−Bradley, and dihedral angle torsion, respectively, E_ω is the improper torsion, and E_{vdW} and E_{Coul} are the van der Waals and Coulombic interactions, respectively. E_{bond} in eq 1 is the total bonding energy, and $E_{nonbond}$ in eq 2 is the total nonbonding energy. Factors expressed in eqs 1 and 2 are all potential-function components in the CHARMM program.

Compactness of the protein structures was monitored by the radius of gyration during the MD simulations. The radius of gyration, $R_g(t)$, was calculated as a function of time using the following equation:

$$R_g(t) = \left[\frac{1}{Na}\sum_{i=1}^{Na}(\mathbf{r}_i(t) - \mathbf{r}_{CM}(t))^2\right]^{1/2} \quad (3)$$

where Na is the number of atoms in a protein molecule, $\mathbf{r}_i(t)$ is the position vector for the ith atom, and $\mathbf{r}_{CM}(t)$ is the center of mass vector of the protein at time t. The center of mass vector is defined by

$$\mathbf{r}_{CM}(t) = \frac{1}{T_{Am}}\sum_{i=1}^{Na}\mathbf{r}_i(t)Am_i \quad (4)$$

where Am_i is the atomic mass of the ith atom in the protein molecule, and T_{Am} is the total atomic mass. Equilibration of a protein structure during MD simulation is monitored using the root-mean-square deviation (rmsd(t)) of the protein. The rmsd(t) is defined as

$$\text{rmsd}(t) = \left[\frac{1}{N}\sum_{i=1}^{N}(r_i(t) - r_i^0)^2\right]^{1/2} \quad (5)$$

where N is the number of atoms being monitored, and $r_i(t)$ and r_i^0 denote the Cartesian coordinates for atom i in the instantaneous structure at time t, and in the reference structure, respectively. The overall flexibility of the native protein structures can also be monitored using the root-mean-square fluctuation (RMSF), defined as follows

$$\text{RMSF} = \left[\frac{1}{N}\sum_{i=1}^{N}(r_i(t) - < r >)^2\right]^{1/2} \quad (6)$$

where N is the number of atoms being monitored, and $r_i(t)$ and $<r>$ denote the Cartesian coordinates for atom i in the instantaneous structure at time t and in the average structure, respectively. Fluctuations were averaged over non-hydrogen atoms for each residue.

The location of charged residues was determined from their solvent accessible surface area (ASA)[45] using a probe radius of 1.4 Å. The percentage of solvent accessibility p (%) was calculated as follows

$$p = \frac{ASA_{X,folded}}{ASA_{X,G-X-G}} \quad (7)$$

where $ASA_{X,folded}$ is the ASA of X residue in the folded state of a given protein, and $ASA_{X,G-X-G}$ is the ASA of the corresponding residue X in tripeptide (Gly−X−Gly) in its the extended state. A residue X in a protein was assigned as being exposed to the solvent if its p value was greater than 20%;[20] otherwise, it was assigned as being buried within the protein. The extended state coordinates ($\phi = \psi = 180°$) were generated using the InsightII program.[46] The steric screening effects provided by the Gly residues neighboring residue X can be modeled, and such steric screening effects reduce ASA values calculated for the X residues in their extended conformations.

Arg, Lys, Asp, and Glu were used for salt bridges (ion pairs) calculations. The number of salt bridges was counted using the criterion that two oppositely charged residues formed a salt bridge if the distance between one Asp or Glu side chain carboxyl oxygen atom and one nitrogen of the positively charged side chain of Arg or Lys was less than 8 Å.[11] Salt bridges formed by His residues were not considered due to the simulated proteins generated with a pH of 7.

The electrostatic solvation free energy, ΔG_{solv}, was calculated using the Poisson−Boltzmann equation (PBEQ) module[47,48] of CHARMM with the atomic radii derived for this type of calculation.[49] The dielectric constants[50] of the aqueous solvent were assigned as 80.0 and 61.8 for 300 K and at

dx.doi.org/10.1021/ci200184y | J. Chem. Inf. Model. 2012, 52, 7−15

350 K, respectively. The dielectric constant for reference environment was 1 and that for the solutes (proteins) was 2.0. The ionic strength was set to 145 mM. The electrostatic potentials were performed using a focusing technique, with a coarse grid spacing of 1.5 Å and a focusing grid spacing of 0.5 Å. An ion exclusion radius (Stern layer) was defined using a probe of 1.4 Å for the dielectric boundary on the basis of a set of atomic radii for proteins. The difference between the electrostatic potential calculated in solution and that calculated in the reference environment yielded the electrostatic contribution to the electrostatic solvation free energy.[48]

RESULTS AND DISCUSSION

Stability of the Simulated Systems. Three MD simulations in explicit solvent were performed for two NMR structures, hyperthermophilic *T. celer* L30e (1GO1),[25] at two different temperatures (300 and 350 K), and its mesophilic homologue, yeast L30e (1CN7),[26] at 300 K. Both *T. celer* L30e and yeast L30e were simulated for 5.0 ns. MD simulations provided the total potential energy of the simulated systems. Total potential energies show fluctuations (eqs 1 and 2) of the simulated systems as functions of time for *T. celer* L30e and yeast L30e at 300 K and for *T. celer* L30e at 350 K, and the energies are approximately constant at both temperatures. The mean values and standard deviations of the total potential energies are −45,201 ± 87 (300 K) and −42,245 ± 144 (350 K) for *T. celer* L30e and −44,608 ± 141 (300 k) for yeast L30e. The standard deviations of the mean values are less than 0.3%.

The compactness of the protein structures, measured using the radius of gyration, was used an indicator of the stability of the native proteins. Figure 2 shows the radii of gyration (R_g),

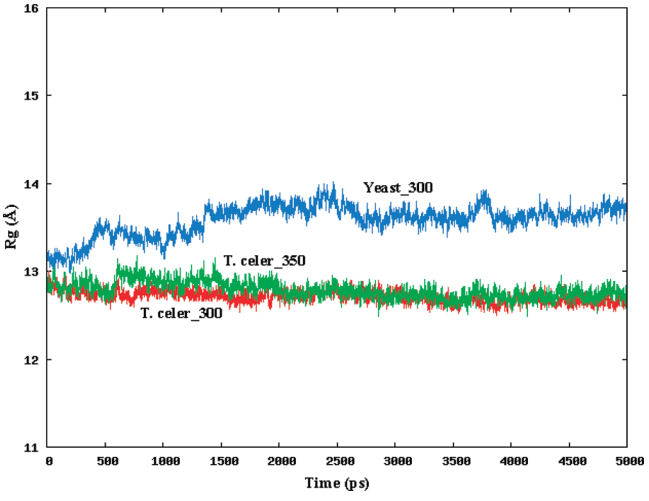


Figure 2. Radius of gyration (R_g) of the proteins as functions of time calculated according to eq 3 for *T. celer* L30e at 300 and 350 K and for yeast L30e at 300 K. Labels for T. celer_300 (red) and T. celer_350 (green) denote the curves for *T. celer* L30e at 300 and 350 K, respectively. Label for Yeast_300 (blue) denotes the curve for yeast L30e at 300 K.

calculated according to eq 3 as functions of time for *T. celer* L30e and yeast L30e at 300 K and *T. celer* L30e at 350 K. As shown in Figure 2, the overall compactness of these curves does not differ significantly, indicating that the native protein structures are sufficiently stable. The average R_g values for *T. celer* L30e are 12.71 ± 0.07 Å and 12.78 ± 0.10 Å at 300 K and at 350 K, respectively, while that for yeast L30e at 300 K is 13.58 ± 0.18 Å.

The fluctuation in protein rmsd over the simulation run is an easy way to check whether the proteins have equilibrated. The equilibration period was determined by the amount of time it took for the rmsd dynamics to fluctuate symmetrically about the proteins most probable trajectory values. Figure 3 shows

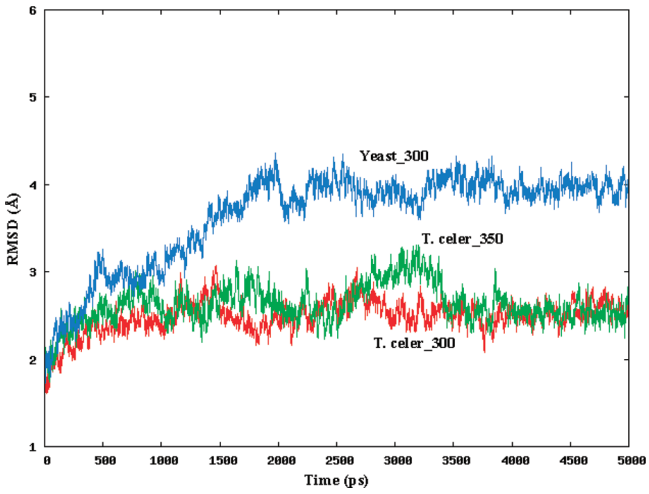


Figure 3. Trajectory of the root-mean-square deviation (rmsd) values calculated according to eq 5 for *T. celer* L30e at 300 and 350 K, and for yeast L30e at 300 K. Labels are the same as in Figure 2.

the rmsd values of the trajectories for *T. celer* L30e and yeast L30e at 300 K and for *T. celer* L30e at 350 K. These rmsd values were calculated with respect to the minimized structures given by to eq 5 and are displayed as functions of time. *T. celer* L30e at 300 and 350 K simulations achieved equilibrium rmsd values in about 1.5 ns, whereas, the yeast L30e 300 K simulation attained their equilibrium in about 2.0 ns. Thus, after 2.0 ns simulation at both temperatures, the trajectories were saved for further analysis. Hereafter, unless otherwise stated, data is computed using the 2−5 ns trajectory periods. Figure 3 shows that the average rmsd values for *T. celer* L30e and yeast L30e at 300 K are 2.36 ± 0.18 Å and 3.64 ± 0.52 Å, respectively, and the rmsd for *T. celer* L30e at 350 K is 2.26 ± 0.22 Å. The average values of rmsd fall between 2.26 and 3.64, indicating that the native protein structures are well maintained at both simulated temperatures and over the simulation periods.

Table 1 shows the secondary structures assignment calculated from the MD simulation trajectories using Kabsch and

Table 1. Contents of Secondary Structures[a] for *T. celer* L30e and Yeast L30e Proteins

		300 K[b]	350 K[c]	NMR[d]	min[e]
T. celer	α-helix	0.43 ± 0.02	0.39 ± 0.03	0.45	0.41
	β-strand	0.17 ± 0.01	0.17 ± 0.02	0.19	0.18
yeast	α-helix	0.37 ± 0.04		0.39	0.35
	β-strand	0.14 ± 0.01		0.19	0.16

[a]Values in this table are computed using the 2−5 ns periods of the trajectories. [b]Average contents of secondary structures calculated for *T. celer* L30e and yeast L30e proteins during the MD simulations at 300 K. [c]Average contents of secondary structures calculated for *T. celer* L30e protein during the MD simulations at 350 K. [d]Contents of secondary structures calculated for *T. celer* L30e (1GO1)[25] and yeast L30e (1CN7)[26] NMR structures. [e]Contents of secondary structures calculated for *T. celer* L30e and for yeast L30e proteins in the final minimized structures.

dx.doi.org/10.1021/ci200184y | J. Chem. Inf. Model. 2012, 52, 7−15

Sander's Dictionary of Protein Secondary Structure (DSSP).[51] For comparison, data calculated from both the NMR determined structures and from the final minimized structures are listed in the last two columns of Table 1. There are no significant differences between the *T. celer* L30e secondary structures at 300 K and at 350 K and yeast L30e at 300 K, as compared to those of the final minimized structures. On the basis of MD trajectory analysis, several characteristic properties, such as total potential energy, radius of gyration, trajectory rmsd, and the secondary structures assignment, indicate that *T. celer* L30e and yeast L30e structures were stable and in equilibrium during the MD simulations.

Flexibility of the Native Protein Structures. The RMSF calculated with respect to the average coordinates of a trajectory and averaged over a simulation run for individual residues is a useful technique to calculate the flexibility of protein structures using MD simulation. Panel (a) of Figure 4 shows the RMSF values of *T. celer* L30e and yeast L30e on a residue basis, at 300 K. Panel (b) of Figure 4 shows the RMSF values for *T. celer* L30e at 300 and 350 K. Each used eq 6. The *T. celer* L30e 3D structures have been determined by both NMR[25] and crystal[32] structures, and both structures are nearly identical, with a rmsd value for the backbone atoms of 0.65 Å.[25] The carboxyl-terminal residues (97−100) are found to be disordered in both structures.[25,32] Also, as shown in panel (c) of Figure 1, the *T. celer* L30e residues 1−96 are well aligned with those of yeast L30e's 8−103 residues. Thus, in panel (a) of Figure 4, the values of RMSF are representative of both the residues 1−96 of *T. celer* L30e and for the residues 8−103 of yeast L30e. Moreover, the yeast L30e residues 8−103 were truncated to 1−96 represented by the *x*-axis in panel (a) of Figure 4. To find the most flexible regions of the proteins, the RMSF values were used to identify local areas of flexibility in the native structures. Panel (a) of Figure 4 shows that there are several well-order residue regions in the *T. celer* L30e structure, while the corresponding residues in yeast L30e are highly disordered. The region between residue 38 and residue 45 in *T. celer* L30e is less flexible compared to the corresponding residues 45−52 in the yeast L30e structure, which are close to the RNA-binding site.[32] The other region (residues 65−80) in the *T. celer* L30e structure contains an α-helix (α4) and a loop connecting α4 to β4. This region is well ordered, whereas the corresponding region (residues 72−87) in the yeast L30e structure shows a highly flexible feature. This region is also close to the RNA-binding site. These results are in agreement with the experimental results of NMR and crystal structures[25,32] of *T. celer* L30e, indicating that these regions are well ordered. However, three residues, Arg76, Pro77, and His78, in the *T. celer* L30e structure are exceptions and are highly flexible, as indicated by solid circles in panel (a) in Figure 4, and these residues have been identified as the worst model in the crystal structure.[32] In the yeast L30e structure, the loop (residues 60−65) connecting α3 to β3 and residues (93−103) at the C-terminus are highly disordered, while these regions correspond to residues 53−58 and to residues 86−96, respectively, in the *T. celer* L30e structure and are well ordered.

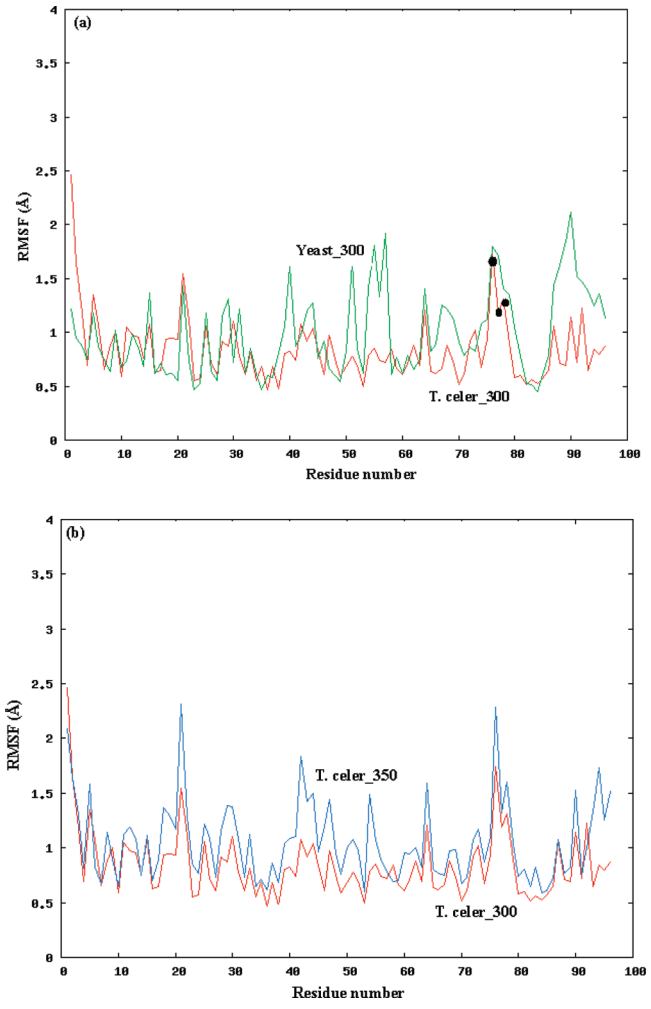


Figure 4. Root mean square fluctuations (RMSFs) per residue calculated according to eq 6 for *T. celer* L30e and for yeast L30e at 300 K (a) and for *T. celer* L30e at 300 and 350 K (b). The values in both (a) and (b) are computed using the 2−5 ns periods of the trajectories. The solid circles in (a) denote the residues (Arg76, Pro77, and His78) in *T. celer* L30e. In (a), labels for T. celer_300 (red) and Yeast_300 (green) in the curves denote the flexibilities for *T. celer* L30e and yeast L30e at 300 K, respectively. In (b), labels for T. celer_300 (red) and T. celer_350 (blue) in the curves denote the flexibilities for *T. celer* L30e at 300 and 350 K, respectively.

As shown in panel (a) of Figure 4, at 300 K the average RMSF value for *T. celer* L30e is 0.85 ± 0.30 Å, and the RMSF value for yeast L30e is 0.99 ± 0.38 Å. As shown in panel (b) of Figure 4, the average *T. celer* L30e RMSF value is 1.06 ± 0.36 Å at 350 K. These results show that the hyperthermophilic *T. celer* L30e is less flexible than its mesophilic homologue, yeast L30e, at 300 K. As shown in panel (b) of Figure 4, the overall values of RMSF for *T. celer* L30e are higher at 350 K than they are 300 K.

Locations of Protein-Charged Residues during MD Simulations. The locations of charged residues are listed in *T. celer* L30e at temperatures 300 and 350 K and in yeast L30e at 300K, along with their solvent accessibility, p (%), calculated using eq 7. *T. celer* L30e shows that nearly all charged residues were exposed to the solvent (p > 20%), except for three residues, Glu6, Lys46, and Glu50, which were buried within the protein during the simulation at 300 K, and one residue, Glu6, was buried when the simulation was carried out at 350 K. Yeast L30e shows that all solvent accessibilities are greater than 20% at 300 K. Thus, all the charged residues in yeast L30e are on the protein surface at 300 K. At pH 7.0, yeast L30e bears a net charge of +8, while *T. celer* L30e has a net charge of +1 because additional ionized acidic group residues exist in *T. celer* L30e. These positive charges on the surface of yeast L30e may produce repulsion potentials,[25] thereby making this protein

Table 2. Distances and Interaction Energies for Ion Pairs[a] in *T. celer* L30e at 300 and 350 K

	300 K			350 K		
ion pair	D (Å)[b]	E_{int} (kcal/mol)[c]	p_{av} (%)[d]	D (Å)[b]	E_{int} (kcal/mol)[c]	p_{av} (%)[d]
R21-D48	6.8 ± 2.2	−50.8 ± 18.6	46.5	9.3 ± 4.6	−45.0 ± 27.2	49.6
R39-E62	5.2 ± 0.9	−99.7 ± 9.3	36.2	5.0 ± 0.8	−101.2 ± 8.1	32.5
R39-E64	4.3 ± 1.0	−67.3 ± 14.0	56.2	4.1 ± 1.1	−82.0 ± 19.5	52.2
R42-D44	4.8 ± 0.2	−110.9 ± 4.1	45.9	5.3 ± 1.4	−99.5 ± 25.0	48.0
R54-E50	4.3 ± 1.0	−92.5 ± 11.9	36.8	5.5 ± 2.2	−71.3 ± 20.0	35.0
R92-E6	6.7 ± 1.0	−52.1 ± 7.6	38.5	5.0 ± 0.4	−117.4 ± 6.5	24.5
K9-E6[e]	**9.7 ± 1.7**	**-43.8 ± 6.8**	**29.7**	3.4 ± 0.9	−116.7 ± 14.4	15.2
K9-D12	7.0 ± 2.2	−48.8 ± 21.9	53.7	10.6 ± 1.0	−27.5 ± 2.0	53.0
K9-E90	4.8 ± 1.5	−71.9 ± 19.4	54.5	7.8 ± 1.7	−43.1 ± 8.8	55.3
K33-D87	6.3 ± 1.6	−60.3 ± 21.5	44.5	7.2 ± 2.2	−53.2 ± 22.5	45.4
K46-E50	3.3 ± 0.7	−105.2 ± 5.4	15.3	3.5 ± 1.0	−101.7 ± 15.7	21.4
K46-E62	4.6 ± 1.4	−90.6 ± 19.7	22.2	5.0 ± 1.6	−83.6 ± 22.9	23.1
K99-E100	6.3 ± 1.5	−110.9 ± 17.9	88.2	7.3 ± 3.2	−139.7 ± 55.8	67.6

[a]Values in this table are computed using the 2−5 ns periods of the trajectories. [b]Distance between one carboxyl oxygen of the negatively charged residue and one side chain nitrogen of the positively charged residue for an ion pair. [c]Interaction energy of an ion pair calculated according to Coulombic interaction energy. [d]Average percentage of solvent accessibility for an ion pair calculated according to eq 7. [e]Distance between this ion pair being 9.7 Å at 300 K, while that of 350 K being 3.4 Å.

unstable at elevated temperatures. On the other hand, *T. celer* L30e has additional acidic residues that can form salt bridge networks[25] on the surface of *T. celer* L30e thus maintaining its thermostability at the higher temperature. Some thermophilic L30e proteins have a lower net positive charge, whereas, some mesophilic L30e proteins have a higher net positive charge, as evidenced[25] by the thermophilic L30e protein from *Pyrococcus horikoshii* and *Methanococcus jannaschii* with net charges of +1 and +3, respectively, and by the mesophilic L30e protein from humans and rice with net charges of +10. Thus, it is likely there is a trend whereby decreasing the net positive charges helps the thermophilic L30e protein family maintain their thermostability at the higher temperatures.

Salt Bridges (Ion Pairs) in the Proteins during MD Simulations. Salt bridge locations in the proteins were evaluated taking the averaged solvent accessibility, p (%), of the two residues forming the salt bridge calculated using eq 7. A salt bridge was designated as being exposed to the solvent if it had an average value of p greater than 20%; otherwise, it was considered as being buried within the protein. Tables 2 and 3 show the salt bridges (ion pairs) with their constituent residues and the distances between the two oppositely charged residues forming the salt bridges, as well as the Coulombic interaction energies between them for *T. celer* L30e and for yeast L30e, respectively. As shown in Table 2, 12 salt bridges are present, and the distances between any pair of oppositely charged residues range from 3.3 to 7.0 Å at 300 K (except for K9-E6) and from 3.4 to 10.6 Å at 350 K. Thus, some of these salt bridges do not maintain distances of less than 8 Å at the higher temperature. Except for the salt bridge K46-E50, with an average solvent accessibility of less than 20%, 11 salt bridges are on the surface of the protein at 300 K. The most striking finding is that the salt bridge (K9-E6) has a mean distance of 9.7 Å at 300 K and approaches a salt bridge formation with a mean distance of 3.4 Å at 350 K. This suggests that this ion pair contributes to the thermal stability of hyperthermophilic *T. celer* L30e at the higher temperature. Two salt bridges (R21-D48 and K9-D12) with mean ion pairs distances of less than 8.0 Å at 300 K fell apart during the simulation at 350 K and separated to longer mean ion pair distances of 9.3 and 10.6 Å, respectively. This implies that these salt bridges do not contribute to the thermal stability of *T. celer* L30e at the higher temperature. One hexad network, R54-E50-K46-E62-R39-E64, is shown in Table 2. The mean distances of the ion pairs that participate in this network fall within the range of 3.3 and 5.2 Å at 300 K and within that of 3.5 and 5.5 Å at 350 K. It is apparent that this network contributes to the thermal stability of *T. celer* L30e at the higher temperature. These results correspond to cluster 2 in Figure 8 of a paper by Wong et al.[25] and represent the tightening of ion pairs[22] at the higher temperature. A triad network, D12-K9-E90, with mean distances of 7.0 Å for D12-K9 and 4.8 Å for K9-E90 at 300 K or E6-K9-E90 with mean distances of 3.4 Å for E6-K9 and of 7.8 Å for K9-E90 at 350 K also can be observed in Table 2. Only the triad, E6-K9-E90, contributes to the thermal stability of *T. celer* L30e at the higher temperature. The triad at both 300 and 350 K corresponds to cluster 1 in Figure 8 of the paper by Wong et al.[25]

Table 3. Distances and Interaction Energies for Ion Pairs[a] in Yeast L30e at 300 K

ion pair	D (Å)[b]	E_{int} (kcal/mol)[c]	p_{av}(%)[d]
R52-E55	3.5 ± 0.3	−109.2 ± 6.1	33.7
K5-E8	5.9 ± 2.3	−71.1 ± 29.2	54.9
K13-D99	3.0 ± 0.5	−117.7 ± 6.6	33.2
K22-E94	6.3 ± 2.7	−68.8 ± 29.4	56.6
K53-E57	4.6 ± 1.7	−91.3 ± 27.5	29.7

[a]Values in this table are computed using the 2−5 ns periods of the trajectories. [b]Distance between one carboxyl oxygen of the negatively charged residue and one side chain nitrogen of the positively charged residue for an ion pair. [c]Interaction energy of an ion pair calculated according to Coulombic interaction energy. [d]Average percentage of solvent accessibility for an ion pair calculated according to eq 7.

As shown in Table 3, there are five salt bridges with ion pair mean distances ranging between 3.0 and 6.3 Å at 300 K. All of these salt bridges are on the protein surface. There are no salt bridge networks present in yeast L30e (Table 3). Thus, on the basis of the data presented in Tables 2 and 3, hyperthermophilic *T. celer* L30e has a greater number of surface salt bridges and salt bridge networks compared to its mesophilic yeast L30e.

Table 4. Electrostatic Solvation Free Energy and the Coulombic Interaction Energy of *T. celer* L30e for Positively Charged Residues[a] at 300 and 350 K

	$\Delta G_{solv}{}^{I}$[b] (kcal/mol)			$E_{Coul}{}^{I}$[d] (kcal/mol)		
residue	300 K	350 K	$\Delta\Delta G_{solv}{}^{I}$[c]	300 K	350 K	$\Delta E_{Coul}{}^{I}$[e]
Arg8	−34.9 ± 4.2	−35.3 ± 4.3	−0.4	−75.4 ± 10.8	−80.4 ± 14.2	−5.0
Lys9	−23.1 ± 13.7	−22.8 ± 11.1	0.3	−119.3 ± 33.4	−129.0 ± 28.8	−9.7
Lys15	−58.2 ± 4.1	−54.7 ± 5.8	3.5	−42.5 ± 12.7	−52.1 ± 16.5	−9.6
Arg21	−61.8 ± 15.1	−64.1 ± 15.6	−2.3	−9.8 ± 33.2	−6.1 ± 39.3	3.7
Lys22	−80.3 ± 4.6	−75.9 ± 4.6	4.4	22.1 ± 10.6	15.0 ± 10.7	−7.1
Lys28	−57.3 ± 4.6	−52.7 ± 5.9	4.6	−36.8 ± 11.0	−45.4 ± 14.2	−8.6
Lys33	−36.5 ± 7.3	−38.4 ± 7.3	−1.9	−85.8 ± 18.7	−84.5 ± 20.4	1.3
Arg39	−1.4 ± 7.6	4.2 ± 8.0	5.6	−176.5 ± 27.4	−201.2 ± 27.3	−24.7
Arg42	−42.3 ± 5.3	−40.6 ± 8.5	1.7	−54.1 ± 22.3	−51.9 ± 29.1	2.2
Lys46	3.4 ± 8.7	6.2 ± 8.3	2.8	−216.2 ± 28.7	−217.3 ± 29.3	−1.1
Arg54	−17.0 ± 5.8	2.8 ± 11.8	19.8	−135.9 ± 16.9	−195.6 ± 39.9	−59.7
Arg76	−61.6 ± 2.8	−58.1 ± 3.6	3.5	−4.7 ± 7.4	−12.5 ± 8.0	−7.8
Arg92	−23.2 ± 9.8	−34.7 ± 5.5	−11.5	−103.4 ± 29.8	−89.6 ± 16.0	13.8
Lys99	−28.8 ± 11.2	−17.8 ± 15.0	11.0	−111.1 ± 33.1	−151.3 ± 44.9	−40.2
<Arg>[f]	−34.6 ± 22.5	−32.3 ± 26.5	2.3	−80.0 ± 63.2	−91.0 ± 78.7	−11.0
<Lys>[f]	−40.1 ± 26.9	−36.6 ± 26.9	3.5	−84.2 ± 74.1	−94.9 ± 76.1	−10.7

[a]Values in this table are computed using the 0−5 ns periods of the trajectories. [b]Electrostatic solvation free energy ($\Delta G_{solv}{}^{I}$) calculated using the PBEQ module[47,48] of CHARMM for a charged residue I. [c]Difference of electrostatic solvation free energies ($\Delta\Delta G_{solv}{}^{I}$) between the mean values at 300 and 350 K. [d]Coulombic interaction energy ($E_{Coul}{}^{I}$) of a charged residue I with the rest of the protein. [e]Difference of the Coulombic interaction energies ($\Delta E_{Coul}{}^{I}$) between the mean values at 300 and 350 K. [f]<Arg> and <Lys> denoted the mean values over all Arg and Lys residues, respectively.

Table 5. Electrostatic Solvation Free Energy and the Coulombic Interaction Energy of *T. celer* L30e for Negatively Charged Residues[a] at 300 and 350 K

	$\Delta G_{solv}{}^{I}$[b] (kcal/mol)			$E_{Coul}{}^{I}$[d] (kcal/mol)		
residue	300 K	350 K	$\Delta\Delta G_{solv}{}^{I}$[c]	300 K	350 K	$\Delta E_{Coul}{}^{I}$[e]
Asp2	−38.4 ± 6.3	−26.2 ± 6.0	12.2	−93.8 ± 11.8	−123.7 ± 12.8	−29.9
Glu6	−20.1 ± 7.2	10.5 ± 14.0	30.6	−128.9 ± 12.9	−241.8 ± 50.2	−112.9
Asp12	−46.2 ± 7.0	−49.1 ± 7.6	−2.9	−91.0 ± 17.0	−82.5 ± 14.6	8.5
Asp44	−0.6 ± 12.6	−14.6 ± 21.2	−14.0	−175.3 ± 34.8	−144.9 ± 53.8	30.4
Glu47	−56.4 ± 6.4	−65.3 ± 5.9	−8.9	−58.3 ± 12.6	−39.9 ± 13.1	18.4
Asp48	−34.8 ± 7.9	−39.7 ± 11.5	−4.9	−95.9 ± 19.0	−85.0 ± 30.0	10.9
Glu50	−18.1 ± 10.6	−30.8 ± 10.2	−12.7	−140.9 ± 34.5	−105.0 ± 25.7	35.9
Glu62	−26.0 ± 10.3	−24.4 ± 8.8	1.6	−132.5 ± 30.0	−137.5 ± 25.0	−5.0
Glu64	−54.3 ± 7.9	−43.5 ± 12.0	10.8	−65.5 ± 18.8	−88.0 ± 29.9	−22.5
Glu69	−34.4 ± 4.8	−32.0 ± 5.4	2.4	−106.8 ± 14.0	−112.2 ± 15.2	−5.4
Asp87	−25.2 ± 8.3	−25.8 ± 9.0	−0.6	−105.5 ± 23.2	−102.0 ± 24.5	3.5
Glu90	−46.6 ± 13.7	−47.9 ± 9.7	−1.3	−98.3 ± 29.8	−93.1 ± 18.1	5.2
Glu100	−120.0 ± 19.5	−97.2 ± 28.0	22.8	−172.6 ± 36.2	−202.6 ± 64.3	−30.0
<Asp>[f]	−29.1 ± 18.0	−31.1 ± 17.2	−2.0	−112.3 ± 39.0	−107.6 ± 39.0	4.7
<Glu>[f]	−47.0 ± 32.8	−33.3 ± 24.0	13.7	−113.0 ± 44.2	−127.5 ± 70.3	−14.5

[a]Values in this table are computed using the 0−5 ns periods of the trajectories. [b]Electrostatic solvation free energy ($\Delta G_{solv}{}^{I}$) calculated using the PBEQ module[47,48] of CHARMM for a charged residue I. [c]Difference of electrostatic solvation free energies ($\Delta\Delta G_{solv}{}^{I}$) between the mean values at 300 and 350 K. [d]Coulombic interaction energy ($E_{Coul}{}^{I}$) of a charged residue I with the rest of the protein. [e]Difference of the Coulombic interaction energies ($\Delta E_{Coul}{}^{I}$) between the mean values at 300 and 350 K. [f]<Asp> and <Glu> denoted the mean values over all Asp and Glu residues, respectively.

PBEQ Calculation for the *T. celer* L30e Protein. To be consistent with the data in Figures 2 and 3, the data involved in this and the following section will be computed using the 0−5 ns periods of the trajectories. The electrostatic solvation free energies, ΔG_{solv}, were calculated using the PBEQ module[47,48] of CHARMM at both 300 and 350 K for *T. celer* L30e. The quantities, −1210.3 ± 89.8 kcal/mol and −1143.9 ± 89.3 kcal/mol, show that the difference in the electrostatic solvation free energies ($\Delta\Delta G_{solv}$) between the mean values at 300 and 350 K is 66.4 kcal/mol. The value of $\Delta\Delta G_{solv}$ for *T. celer* L30e is an unfavorable value of 66.4 kcal/mol, suggesting that *T. celer* L30e is less solvated at the higher temperature. This result is a consequence of the decrease in magnitude of the desolvation penalty with increasing temperature.[22]

Electrostatic Solvation Free Energy and Coulombic Interaction Energy for Individual Charged Residues. The values of the electrostatic solvation free energy, $\Delta G_{solv}{}^{I}$, and those of Coulombic interaction energy, $E_{Coul}{}^{I}$, evaluated for a charged residue I in *T. celer* L30e during the simulation at 300 and 350 K are entered in Table 4 for positively charged residues and Table 5

dx.doi.org/10.1021/ci200184y | J. Chem. Inf. Model. 2012, 52, 7−15

for negatively charged residues. The values of ΔG_{solv}^{I} were calculated using the PBEQ module[47,48] of CHARMM for a charged residue I. Values for E_{Coul}^{I} were calculated according to Coulombic interactions for a charged residue I along with the rest of the protein. The sum of ΔG_{solv}^{I} for all charged residues is −1044.4 kcal/mol at 300 K and −904.2 kcal/mol at 350 K. As compared to the PBEQ calculation for the *T. celer* L30e electrostatic solvation free energies described in the previous section, these values represent 86% and 79% of the total electrostatic solvation free energy (ΔG_{solv}) at 300 and 350 K, respectively. This shows the charged residues provide the major contribution to the total electrostatic solvation free energy of *T. celer* L30e.

As shown in Tables 4 and 5, with the exception of Lys46 at 300 K and Arg39, Lys46, Arg54, and Glu6 at 350 K, the values of ΔG_{solv}^{I} are negative for all charged residues. At 300 K, the average electrostatic solvation free energies for positively charged residues are −34.6 kcal/mol (Arg) and −40.1 kcal/mol (Lys), and mean values for the negatively charged residues are −29.1 kcal/mol (Asp) and −47.0 kcal/mol (Glu). There is a similar trend observed for the average electrostatic solvation free energies of these four residue types at 350 K. The Lys residue is more effectively solvated than the Arg residue, and it is likely that the Lys residue has a more concentrated charge on its side chain. The Glu residue is better solvated than the Asp residue, possibly because the Glu residue has a longer side chain. As shown in Tables 4 and 5, all the values of E_{Coul}^{I} are negative, except that for Lys22. By examining the values of $\Delta\Delta G_{solv}^{I}$ and ΔE_{Coul}^{I}, we note that the values have opposite signs for each charged residue, except for Arg8 and Arg42. This reveals a negative correlation with a Pearson correlation coefficient[52] of −0.93, a slope of −2.65, and an intercept of −1.62 kcal/mol between the values of $\Delta\Delta G_{solv}^{I}$ and ΔE_{Coul}^{I}. The average values of $\Delta\Delta G_{solv}^{I}$ indicate that Arg, Lys, and Glu residues are less solvated at the higher temperature. The average values of ΔE_{Coul}^{I} reveal that the same residues (Arg, Lys, and Glu) have stronger Coulombic interactions with other residues (other charged and neutral residues) at the higher temperature. Both could be caused by decreased solvent density at the higher temperature. Thus, it is interesting to find that a charged residue with a large and positive (or negative) value for $\Delta\Delta G_{solv}^{I}$ and a large and negative (or positive) value for ΔE_{Coul}^{I} should be stable at the higher temperature. This finding may be useful as a guide to designing side-directed mutagenesis experiments on *T. celer* L30e to disrupt the Coulombic interactions between ion pairs. For the mutation of Glu6 by alanine (E6A), the E6A variant exhibits a large decrease in T_m (ΔT_m = −8.6 °C)[53] because the value of $\Delta\Delta G_{solv}^{I}$ is 30.6 kcal/mol, while that of ΔE_{Coul}^{I} is −112.9 kcal/mol (Table 5). The mutation of Arg92 by alanine or by methionine (R92A or R92M) gives a moderate decrease in T_m (ΔT_m = −3 °C)[53] because the value of $\Delta\Delta G_{solv}^{I}$ is −11.5 kcal/mol and that of ΔE_{Coul}^{I} is 13.8 kcal/mol (Table 4). For the mutation of Glu90 by alanine (E90A), the E90A variant gives only a small decrease of T_m (ΔT_m = −0.9 °C)[53] because the value of $\Delta\Delta G_{solv}^{I}$ is −1.3 kcal/mol, and that of ΔE_{Coul}^{I} is 5.2 kcal/mol for Glu90 (Table 5). These three examples from the experimental works of Lee et al.[53] may reinforce the results of this study.

CONCLUSIONS

MD simulations were performed for the hyperthermophilic *T. celer* L30e at 300 and 350 K and for its mesophilic counterpart, yeast L30e, at 300 K in explicit solvent for 5.0 ns. The simulations were conducted with the aid of the CHARMM program. The stability and equilibration of the native protein structures were monitored by calculating several characteristic properties. During 300 K simulation, yeast L30e exhibited greater flexibility than its hyperthermophilic homologue, *T. celer* L30e, although *T. celer* L30e was more flexible at 350 K than it was at 300 K. The locations of charged residues in the *T. celer* L30e protein were determined from their solvent accessibility, *p* (%), during the MD simulations at 300 and 350 K. We found that indicated that 89% of charged residues exist on the surface of *T. celer* L30e at 300 K, and 96% were present on surface at 350 K. Some of the charged residues present on the *T. celer* L30e surface form salt bridge networks, which maintain the thermostability of this protein at elevated temperatures because of its net charge of +1. In contrast, 100% of charged residues are on the surface of yeast L30e. These charged residues produce charge repulsion potentials due to its net charge of +8, which destabilizes the protein. For charged residue I in *T. celer* L30e, the values of $\Delta\Delta G_{solv}^{I}$ and ΔE_{Coul}^{I} have opposite signs. The value of the Pearson correlation coefficient is large and negative. It is interesting to note that a given charged residue with a large positive (or negative) $\Delta\Delta G_{solv}^{I}$ term and a large negative (or positive) ΔE_{Coul}^{I} term is stabilized at the higher temperature. This finding may prove useful in side-directed mutagenesis experiments on *T. celer* L30e for disruption of Coulombic interactions.

AUTHOR INFORMATION

Corresponding Author

*Phone: +886-4-2339-8316. Fax: +886-4-2332-1019. E-mail: kjlee@asia.edu.tw.

ACKNOWLEDGMENTS

The National Center for High-Performance Computing in Taiwan is gratefully acknowledged for providing computer facilities. Ms. Chuan-Mei Lee of Harvard University has helped with the English editing of this manuscript.

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

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

(4) 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.

(5) 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.

(6) 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.

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

(8) Haney, P.; Konisky, J.; Koretke, K. K.; Luthey-Schulten, Z.; Wolynes, P. G. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. *Proteins* **1997**, *28*, 117−130.

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

(10) Kumar, S.; Tsai, C.-J.; Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **2000**, *13*, 179−191.
(11) Szilágyi, A.; Závodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. *Structure* **2000**, *8*, 493−504.
(12) Chakravarty, S.; Varadarajan, R. Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study. *Biochemistry* **2002**, *41*, 8152−8161.
(13) Gianese, G.; Bossa, F.; Pascarella, S. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins* **2002**, *47*, 236−249.
(14) Sadeghi, M.; Naderi-Manesh, H.; Zarrabi, M.; Ranjbar, B. Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **2006**, *119*, 256−270.
(15) Greaves, R. B.; Warwicker, J. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. *BMC Struct. Biol.* **2007**, *7*, 18−40.
(16) Paiardini, A.; Sali, R.; Bossa, F.; Pascarella, S. "Hot cores" in proteins: Comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms. *BMC Struct. Biol.* **2008**, *8*, 14−31.
(17) Maugini, E.; Tronelli, D.; Bossa, F.; Pascarella, S. Structural adaptation of the subunit interface of oligomeric thermophilic and hyperthermophilic enzymes. *Comp. Biol. Chem.* **2009**, *33*, 137−148.
(18) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(19) Cambillau, C.; Claverie, J.-M. Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **2000**, *275*, 32383−32386.
(20) 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.
(21) 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.
(22) de Bakker, P. I. W.; Hünenberger, P. H.; McCammon, J. A. Molecular dynamics simulations of the hyperthermophilic protein Sac7d from *Sulfolobus acidocaldarius*: Contribution of salt bridges to thermostability. *J. Mol. Biol.* **1999**, *285*, 1811−1830.
(23) Bae, E.; Philips, 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.
(24) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(25) Wong, K.-B.; Lee, C.-F.; Chan, S.-H.; Leung, T.-Y.; Chen, Y. W.; Bycroft, M. Solution structure and thermol stability of ribosomal protein L30e from hyperthermophilic archaeon *Thermococcus celer*. *Protein Sci.* **2003**, *12*, 1483−1495.
(26) Mao, H.; Williamson, J. R. Local folding coupled to RNA binding in the yeast ribosomal protein L30. *J. Mol. Biol.* **1999**, *292*, 345−359.
(27) Dabeva, M. D.; Warner, J. R. The yeast ribosomal protein L32 and its gene. *J. Biol. Chem.* **1987**, *262*, 16055−16059.
(28) Mager, W. H.; et al. A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **1997**, *25*, 4872−4875.
(29) Eng, F. J.; Warner, J. R. Structural basis for the regulation of splicing of a yeast messenger RNA. *Cell* **1991**, *65*, 797−804.
(30) Li, B.; Vilardell, J.; Warner, J. R. An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1596−1600.
(31) Vilardell, J.; Warner, J. R. Ribosomal protein L32 of *Saccharomyces cerevisiae* influences both the splicing of its own transcript and the processing of rRNA. *Mol. Cell. Biol.* **1997**, *17*, 1959−1965.
(32) Chen, Y. W.; Bycroft, M.; Wong, K.-B. Crystal structure of ribosomal protein L30e from the extreme thermophile *Thermococcus celer*: Thermal stability and RNA binding. *Biochemistry* **2003**, *42*, 2857−2865.
(33) Lazaridis, T.; Lee, I.; Karplus, M. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **1997**, *6*, 2589−2605.
(34) Colombo, G.; Merz, K. Jr. Stability and activity of mesophilic subtilisin E and its thermophilic homolog: Insights from molecular dynamics simulations. *J. Am. Chem. Soc.* **1999**, *121*, 6895−6903.
(35) 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.
(36) Tang, L.; Liu, H. A comparative molecular dynamics study of thermophilic and mesophilic ribonulease HI enzymes. *J. Biomol. Struct. Dyn.* **2007**, *24*, 379−392.
(37) Basu, S.; Sen, S. Turning a mesophilic protein into thermophilic one: A computational approach based on 3D structural features. *J. Chem. Inf. Model* **2009**, *49*, 1741−1750.
(38) 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.
(39) MacKerell, A. D. Jr.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(40) 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*, 926−935.
(41) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(42) 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.
(43) York, D. M.; Darden, T. A.; Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(44) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667−683.
(45) Lee, B. K.; Richards, F. M. The interpretation of protein structures. Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(46) Accelrys, S. I. *InsightII*, version 2000.1; San Diego, CA, 2000.
(47) Im, W.; Beglov, D.; Roux, B. Continnum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Bolzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(48) Jo, S.; Vargyas, M.; Vasko-Szedlar, J.; Roux, B.; Im, W. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* **2008**, *36*, W270−W275.
(49) Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics based on molecular free energy simulations. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.
(50) *CRC Handbook of Chemistry and Physics*, 89th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2008.
(51) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.
(52) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in FORTRAN*; Cambridge University Press: New York, 1986.
(53) Lee, C.-F.; Allen, M. D.; Bycroft, M.; Wong, K.-B. Electrostatic interactions contribute to reduced heat capacity change of unfoldig in a thermophilic ribosomal protein L30e. *J. Mol. Biol.* **2005**, *348*, 419−431.