

## RESEARCH ARTICLE

# Thermostabilization mechanisms in thermophilic versus mesophilic three-helix bundle proteins

Catrina Nguyen | Lauren M. Yearwood | Michelle E. McCully Department of Biology, Santa Clara University,  
Santa Clara, California, USA**Correspondence**Michelle E. McCully, Department of Biology,  
Santa Clara University, 500 El Camino Real,  
Santa Clara, CA 95053, USA.  
Email: memccully@scu.edu**Funding information**National Institute of General Medical Sciences,  
Grant/Award Number: GM134439**Abstract**

The engineered three-helix bundle, UVF, is thermostabilized entropically due to heightened, native-state dynamics. However, it is unclear whether this thermostabilization strategy is observed in natural proteins from thermophiles. We performed all-atom, explicit solvent molecular dynamics simulations of two three-helix bundles from thermophilic *H. butylicus* (2lvsN and 2lvsC) and compared their dynamics to a mesophilic three-helix bundle, the Engrailed homeodomain (EnHD). Like UVF, 2lvsC had heightened native dynamics, which it maintained without unfolding at 100°C. Shortening and rigidification of loops in 2lvsN and 2lvsC and increased surface hydrogen bonds in 2lvsN were observed, as is common in thermophilic proteins. A buried disulfide and salt bridge in 2lvsN and 2lvsC, respectively, provided some stabilization, and addition of a homologous disulfide bond in EnHD slowed unfolding. The transferability and commonality of stabilization strategies among members of the three-helix bundle fold suggest that these strategies may be general and deployable in designing thermostable proteins.

**KEYWORDS**

molecular dynamics, protein dynamics, protein engineering, protein thermostability, thermophile

## 1 | INTRODUCTION

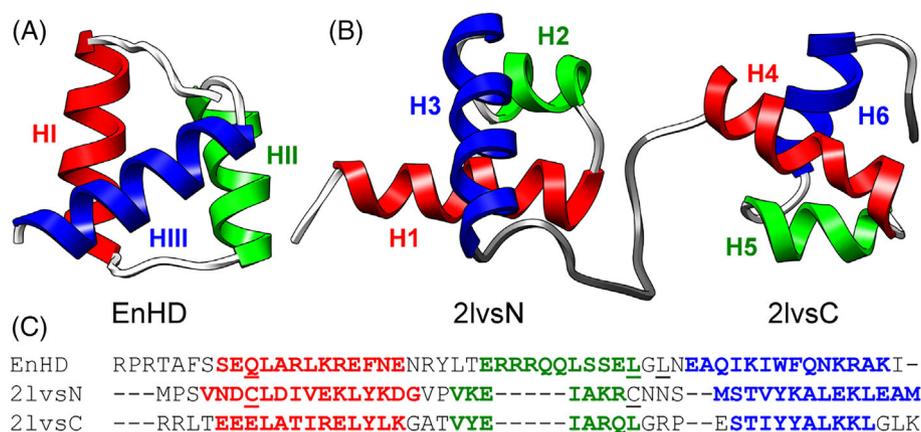
De novo designed proteins tend to be highly stable and thermostable, if they fold at all.<sup>1–3</sup> Computationally and rationally designed proteins often include structural features that are associated with stability, although few projects explicitly attempt to design highly thermostable proteins.<sup>4,5</sup> It remains an open question whether the strategies that make designed proteins thermostable in practice are the same as those found in nature, in the proteins of thermophilic organisms. Furthermore, if the strategies that designed proteins use are different from those employed by nature, will they be compatible with function, the ultimate goal of protein engineering?

We previously explored a pair of three-helix bundle proteins, EnHD and UVF. The Engrailed homeodomain (EnHD, Figure 1A) is a transcription factor from *D. melanogaster* that has ultra-fast folding and unfolding kinetics.<sup>6</sup> UVF was *de-novo* designed by the Mayo Group based on EnHD's backbone, and it is highly thermostable ( $T_m > 99^\circ\text{C}$ ).<sup>7</sup> Based on molecular dynamics (MD) and coarse-grained simulations, UVF was hypothesized to be thermostabilized enthalpically due to surface salt bridges and entropically due to the nanosecond-timescale dynamics of its backbone and buried hydrophobic side chains.<sup>8</sup> While surface salt bridges are common and contribute to thermostability among naturally thermostable proteins,<sup>9–13</sup> entropic thermostabilization due to heightened dynamics has been less-well documented.<sup>14–16</sup> As UVF does not perform any function, it is unclear whether this method of thermostabilization is compatible with

Catrina Nguyen and Lauren M. Yearwood contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.



**FIGURE 1** Protein structures and sequences for EnHD and 2lvs. (A) Mesophilic EnHD (PDB 1enh, residues 3–56) colored by helix (HI 10–22, HII 28–38, HIII 42–55). (B) Thermophilic Cbp2<sub>Hb</sub>/2lvs (PDB 2lvs, residues 1–105) colored by helix (2lvsN residues 1–45, H1 4–18, H2 21–27, H3 32–45; 2lvsC residues 59–103, H4 63–75, H5 79–86, H6 91–100) showing excluded linker regions (residues 46–59, 104–105). (C) Aligned sequences with helical regions indicated in bold and residues of interested underlined (EnHD Q12, L38, L40; 2lvsN C7, C28; 2lvsC E65, P/T68, R88)

function and therefore could have broad applicability in designing thermostable proteins. However, if thermostabilization via dynamics is observed in a similar, naturally occurring, thermophilic protein, it would provide evidence that this strategy is not only compatible with function but may also be an untapped method for designing thermostable proteins. Here, we set out to investigate whether a homologous, naturally occurring thermophilic three-helix bundle was thermostabilized similarly to UVF.

*Hyperthermus butylicus* is a thermophilic archaeobacterium isolated off the coast of the Portuguese island of São Miguel. Its optimum growth temperature is between 95 and 106°C, suggesting that its proteins are highly thermostable.<sup>17</sup> It has a class 2 CRISPR DNA repeat-binding protein with two three-helix-bundle domains (Cbp2<sub>Hb</sub>, Figure 1B), which bind DNA.<sup>18</sup> The two bundles have 6 and 22% sequence identity to EnHD, respectively (Figure 1C), and we will refer to them here by PDB code, noting them as N- and C-terminal, as 2lvsN and 2lvsC. We used room- and high-temperature all-atom, explicit solvent MD simulations to investigate the structural and dynamic properties that led to these three-helix bundle proteins' high thermostability, and we compared their thermostabilization strategies with UVF's.

Like UVF, the C-terminal bundle (2lvsC) had heightened backbone dynamics, although its core side chains did not; the effect was present but more subtle for the N-terminal bundle (2lvsN). Also like UVF, 2lvsN had many surface polar interactions, but 2lvsC had fewer than EnHD. Both domains had a shorter and more-rigid first loop than EnHD, as is common in thermostable proteins. But unexpectedly, 2lvsC and 2lvsN buried less hydrophobic surface area than EnHD upon folding. Both domains contain a strong core interaction; 2lvsN has a disulfide bond and 2lvsC has a salt bridge. These two interactions contributed to but did not fully explain the proteins' extreme thermostabilities. Insertion of a disulfide bond or salt bridge in an equivalent position in EnHD was somewhat thermostabilizing, although the disulfide bond was more effective.

## 2 | METHODS

### 2.1 | Preparation of protein structures

Starting structures were retrieved from the Protein Data Bank (PDB) for EnHD (PDB 1enh)<sup>19</sup> and 2lvs (PDB 2lvs).<sup>18</sup> For the 2lvs NMR structure, model 1 was chosen, and the structure was divided into its two three-helix-bundle domains, 2lvsN (residues 1–45) and 2lvsC (residues 59–103). In 2lvsN, the two cysteine residues (residues 7 and 28) had been mutated to serine for ease of structure determination, so they were mutated back to cysteine by manually renaming the residues and the OG atoms to SG. To create the disulfide-bonded version (2lvsN-ss), the rotamers of both cysteines were chosen such that the side chains pointed toward each other in Chimera using the Dunbrak rotamer library.<sup>20,21</sup> The NMR structure of 2lvs contains a threonine at residue 68, whereas the Uniprot sequence (A2BLH2) contains a proline. As it was unclear which was biologically accurate, the 2lvsC<sub>P68</sub> construct was also built in Chimera. To build the salt bridge and disulfide bond mutations in 2lvsC and EnHD (2lvsC<sub>T68</sub>-nosb, 2lvsC<sub>P68</sub>-nosb, EnHD-ss, and EnHD-sb), rotamers were again assigned using Chimera. Helical content in the experimental structures was assessed using Chimera's implementation of the DSSP algorithm.

### 2.2 | Molecular dynamics simulations

All protein structure files were created in VMD<sup>22</sup> using autopsf. A DISU patch was added during the autopsf step for 2lvsN-ss and EnHD-ss. Hydrogen atoms were added consistent with neutral pH using autopsf, and none of the proteins contains a histidine. Side chain protonation states were unchanged at high temperature, as their pK<sub>a</sub>s are not expected to change significantly.<sup>23,24</sup> Using NAMD 2.11<sup>25</sup> with the CHARMM36m force field,<sup>26–28</sup> 1000 steps of conjugate gradient minimization was performed for each structure. The proteins

were solvated in a cubic box of TIP3P water<sup>29</sup> with edge lengths of either 46 or 50 Å, and KCl ions<sup>30</sup> were added to neutralize the system at a concentration of 150 mM. The full system was minimized for an additional 100 steps then heated to 25 or 100°C and equilibrated for 5 ps with a 2-fs time step. The NPT ensemble (constant number of particles, pressure, and temperature) was maintained using a Langevin thermostat and barostat, and bonds with hydrogen atoms were constrained using SHAKE. Full-system, periodic electrostatics were calculated using particle mesh Ewald, and a smooth cutoff of 8 Å was applied to van der Waals interactions. Five 100-ns production runs were performed for each protein at both temperatures with structures saved every 1 ps. The simulations of wild type EnHD have been previously described.<sup>8</sup>

### 2.3 | Simulation analysis

Analysis of the coordinate files was performed using the *in lucem* molecular mechanics (*ilmm*) software suite.<sup>31</sup> The C $\alpha$  RMSD was calculated relative to the minimized starting structure over the helical core: residues 10–54 for EnHD, 3–43 for 2lvsN, and 63–100 for 2lvsC; and a 100-ps rolling average was plotted. In addition, the core C $\alpha$  RMSD was calculated between all structures from all 10 simulations at a 10-ps resolution in a pairwise manner, resulting in an all-versus-all core C $\alpha$  RMSD matrix.

Solvent accessible surface area was quantified for each residue using *ilmm*'s implementation of the Lee and Richards algorithm<sup>32</sup> and broken down by buried and surface residues. Buried residues were defined as those in the minimized starting structure that had  $\leq 30\%$  of the SASA of the same residue in an extended pentapeptide, Gly-Gly-Xxx-Gly-Gly. Buried residues for each protein were assigned as: 8, 12, 16, 19, 20, 26, 30, 31, 34, 35, 38, 40, 42, 44, 45, 48, and 49 for EnHD; 11, 12, 14, 19, 21, 24, 25, 28, 35, 38, 39, and 42 for 2lvsN; and 61, 65, 69, 77, 82, 83, 86, 88, 92, 93, and 96 for 2lvsC. SASA was also broken into hydrophobic and polar based on atomic partial charges.

Contacts between residue pairs were counted if they contained at least one pair of atoms participating in a hydrogen bond, hydrophobic interaction, or “other” interaction. Three atoms made a hydrogen bond if the donor-hydrogen-acceptor angle was within 45° of linearity and the donor and acceptor atoms were  $< 2.6$  Å apart. Two carbons made a hydrophobic contact if they were bound to at least one hydrogen atom and were  $< 5.4$  Å apart. Any two other non-hydrogen atoms that were  $< 4.6$  Å apart were defined as an “other” contact. Subsets of contacts that occurred between two buried residues or two surface residues were considered. Salt bridges were counted if the side-chain oxygen atoms of aspartate and glutamate and the nitrogen and associated hydrogen atoms of arginine and lysine made a hydrogen bond or “other” interaction.

Helical content was assessed using *ilmm*'s implementation of the DSSP algorithm<sup>33</sup> as well as based on backbone  $\phi/\psi$  angles. Those residues that were classified as  $\alpha$ -helix by DSSP  $> 25\%$  of the time in simulations at 25°C were considered helical, and others were considered loops.

Transition states were identified for EnHD, EnHD-ss, and EnHD-sb in each of the five simulations at 100°C where the protein

denatured using the method developed by Li and Daggett.<sup>34,35</sup> Briefly, all-versus-all core C $\alpha$  RMSD matrices were calculated for each simulation at 100°C with a 10-ps resolution. The exit of the native cluster was identified as the transition state on a three-dimensional projection of a multidimensional scaling of the matrix in R.<sup>36</sup> These times were confirmed by visual inspection of the trajectories, using movement of H111 away from the H1-H11 core as the first step of unfolding.<sup>37</sup>

## 3 | RESULTS AND DISCUSSION

Here we will describe some general features of our MD simulations of EnHD, 2lvsN, and 2lvsC. Next, we will discuss several structural and dynamic features common to thermophilic proteins and observe whether 2lvs shared these features. We will investigate the backbone and side chain dynamics of 2lvs to determine whether it was thermostable for a similar reason as UVF. Finally, we will determine the contribution of the core disulfide bond from 2lvsN and salt bridge from 2lvsC to their thermostabilization and observe whether a similar interaction engineered into EnHD was stabilizing.

### 3.1 | MD simulations of EnHD reproduced the stability in EnHD at 25 and 100°C

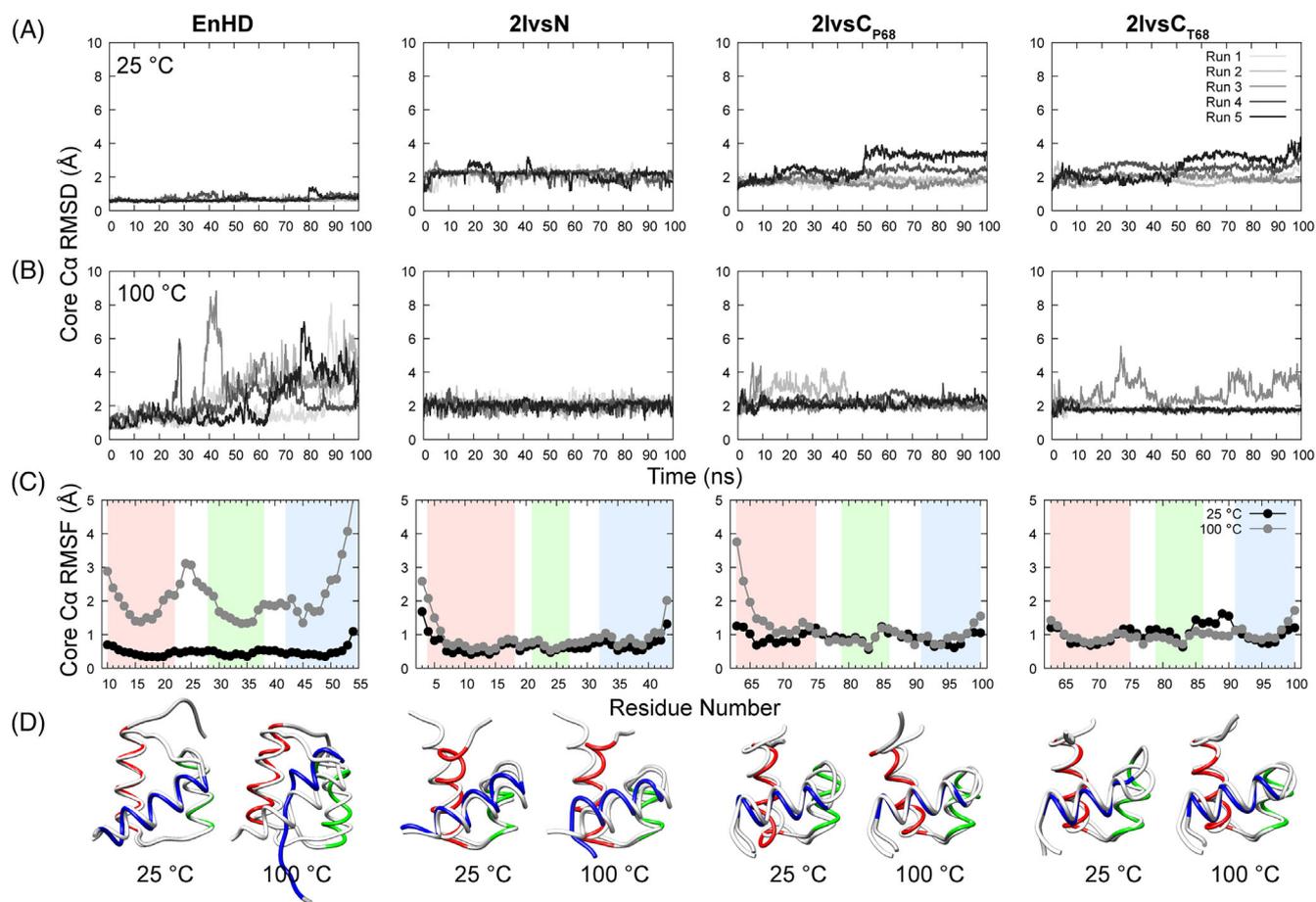
The EnHD simulations have been described previously,<sup>8</sup> but briefly, the protein had completely different dynamics at 100°C compared to 25°C, evidenced by its increasing core C $\alpha$  RMSD over time at 100°C and significantly higher core C $\alpha$  RMSF at 100°C than 25°C (Figure 2A–C and Figure S1). EnHD behaved as expected, compared with experimental results, showing stability at 25°C and the early steps of unfolding at 100°C (Figure 2D).

### 3.2 | Bending of H1 in 2lvsN

H1 of 2lvsN was straight in the initial, NMR structure, but it bent toward the core over the simulation time, with the frequent loss of the backbone hydrogen bond between Asp6 and Ile10 (Figure 2D). This bending allowed additional burial of Val4 and Leu8 with Thr34, Lys37, Ala38, Leu42, and Met45. There was varying but minimal bending in H1 in many of the NMR models, so it is unclear why this bending was more pronounced and so common in our MD simulations.

### 3.3 | The ambiguous amino acid at position 68 of 2lvsC

2lvsC contains a threonine at residue 68 in its PDB structure but a proline in its Uniprot sequence. It was unclear which residue is accurate, so we performed simulations with both, 2lvsC<sub>P68</sub> and 2lvsC<sub>T68</sub>. H4, which contains residue 68, was far more stable at 100°C in



**FIGURE 2** Thermostability and heightened backbone dynamics in 2lvs. (A)  $C\alpha$  RMSD measured over the helical core residues (EnHD 10–54, 2lvsN 3–43, 2lvsC 63–100) and averaged over a 100-ps window at for five simulations at 25°C and (B) 100°C. (C) Average  $C\alpha$  RMSF for the helical core residues at 25°C and 100°C,  $n = 5$ . (D) Representative final MD structures (colored by helix) compared to the minimized starting structure (white)

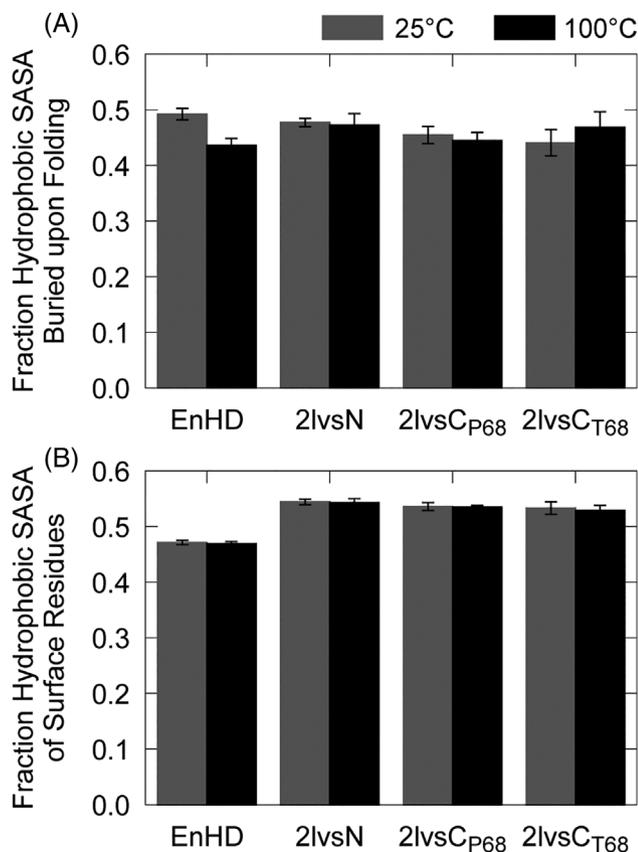
2lvs<sub>C<sub>T68</sub></sub> than 2lvs<sub>C<sub>P68</sub></sub>, based on both DSSP and  $\phi/\psi$  definitions for  $\alpha$ -helix (Figure S2). In all simulations of 2lvs<sub>C<sub>P68</sub></sub> at 100°C, H4 unwound at the N-terminus due to the inability of Glu64 and Pro68 to form a main-chain hydrogen bond. However, even in these simulations, the hydrophobic core and overall three-helix bundle fold over residues 67–100 was maintained. Curiously, 2lvs<sub>C<sub>T68</sub></sub> had a lower core  $C\alpha$  RMSD at 100°C than at 25°C (Figure 2A,B and Figure S1) in four of the five replicates, potentially showing higher stability at the higher temperature. Prolines are not usually found in helices, and they are even less common among helices in thermophilic proteins.<sup>11</sup> It is our suspicion that there was an error at this position in the sequence deposited in Uniprot, and the sequence that was cloned for structural characterization containing Thr68 is the correct wild type sequence.

### 3.4 | The role of secondary structure and hydrophobic surface area in 2lvs thermostability

Thermophilic proteins tend to have shorter loops and higher proportions of their residues in  $\alpha$ -helices than mesophilic proteins.<sup>10,11</sup> It is difficult to quantify the fraction of  $\alpha$ -helix in 2lvs due to the

somewhat arbitrary definition of the N- and C-termini of the domains. However, it is clear that the loop between the first two helices is shorter in 2lvsN and 2lvsC than in EnHD, and this difference is maintained during the MD simulations at 25°C (Table S1). The first loop is also more rigid, relative to the flexibility of the helices, in 2lvs than in EnHD (Figure 2C). Thermophilic proteins often have shorter and more rigid loops than mesophilic proteins.<sup>10,11</sup>

Burial of hydrophobic solvent-accessible surface area (SASA) is a driving force in protein folding, and it is particularly important for the extreme stability found in both thermophilic and designed proteins.<sup>1,10–12,38–40</sup> On the contrary, all three thermophilic structures buried a smaller fraction of the hydrophobic surface area compared to their extended structures than EnHD did, particularly 2lvsC (Figure 3A). 2lvsC had more bulky residues with hydrophobic content (e.g., tyrosine) on its surface. It is unclear whether this feature is thermostabilizing for 2lvsC or provides a potential avenue for further thermostabilization. Similarly, in the 2lvs proteins, a higher fraction of the SASA of their surface residues was hydrophobic compared to EnHD (Figure 3B). As the hydrophobic effect is weaker at high temperature,<sup>41</sup> it may be less unfavorable for a higher proportion of 2lvs's surface area to be hydrophobic, compared to EnHD.



**FIGURE 3** Role of hydrophobic SASA in the thermostability of 2lvs. (A) Average hydrophobic surface area in MD simulation divided by the hydrophobic surface area in the extended structure. (B) Average hydrophobic surface area of the surface residues divided by total surface area of the surface residues. Average  $\pm$  SD,  $n = 5$

### 3.5 | The role of backbone dynamics in 2lvs thermostability

In previous work on UVF, an engineered, thermostable, three-helix-bundle protein, we found that heightened backbone dynamics provided structural evidence of entropic thermostabilization.<sup>8</sup> Like UVF, 2lvsN was somewhat more dynamic at 25°C than EnHD and maintained these heightened dynamics at 100°C without unfolding (Figure 2 and Figure S1). 2lvsC was even more dynamic than 2lvsN at both 25 and 100°C (Figure 2C and Figure S1), but again maintained its folded structure at both temperatures. This observation provides initial evidence that stabilization via dynamics may be found in naturally thermostable proteins and could be a more general method of thermostabilization.

### 3.6 | Contact patterns underlying thermostability

UVF's heightened backbone dynamics are facilitated by highly mobile side chains within the fully-hydrophobic core.<sup>8,42</sup> 2lvsN and 2lvsC, like UVF, had increased backbone dynamics at 25°C, relative to EnHD, and they maintained these heightened dynamics without unfolding at

100°C. Therefore, we might expect the pattern of increased side chain dynamics observed in UVF to be present in 2lvsN and 2lvsC as well.

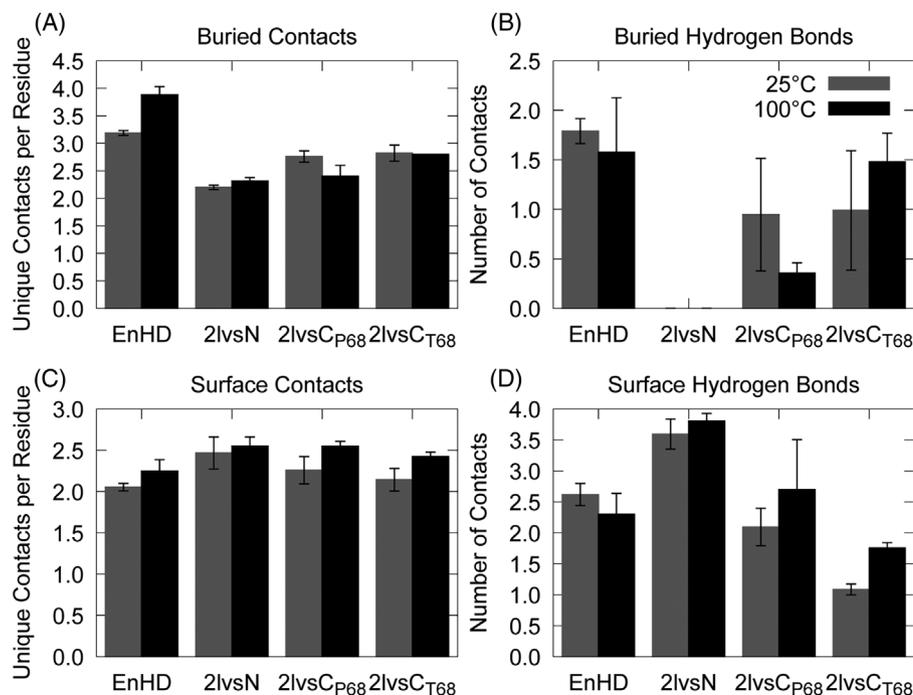
The heightened side chain dynamics in UVF's fully hydrophobic core led to an increase in the unique pairs of contacts made by buried residues, relative to EnHD. While 2lvsN, like UVF, had no buried hydrogen bonds, (Figure 4B), it had fewer unique buried contacts than EnHD, indicative of rigid core packing (Figure 4A). All 16 of UVF's buried residues are hydrophobic, whereas only 11 of 2lvsN's 12 buried residues are. Glu12, the single non-hydrophobic buried residue, is oriented such that its carboxyl group is solvent accessible. As side chain interactions in 2lvsN's core were not as dynamic as in UVF's, it may be that thermostabilization via dynamics requires all—not most—buried residues to be hydrophobic. Perhaps Glu12 played a critical role in rigidifying 2lvsN's core or attracting hydrogen-bonding interactions from surface residues, and likely, 2lvsN's core-to-surface disulfide bond did as well.

Like 2lvsN, 2lvsC had fewer unique core contacts than EnHD (Figure 4A), but unlike UVF and 2lvsN, 2lvsC did have a buried hydrogen bond (Figure 4B). The most frequent core hydrogen bond was observed between Glu65 and Arg88, with an occupancy of about 65%. This long-range hydrogen bond connects H4 and the H5-H6 loop in a homologous location as 2lvsN's disulfide bond between Cys7 (H1) and Cys28 (H2-H3 loop) (Figure 5B). It may be that these strong interactions contributed to side-chain rigidification among buried residues of 2lvsN and 2lvsC.

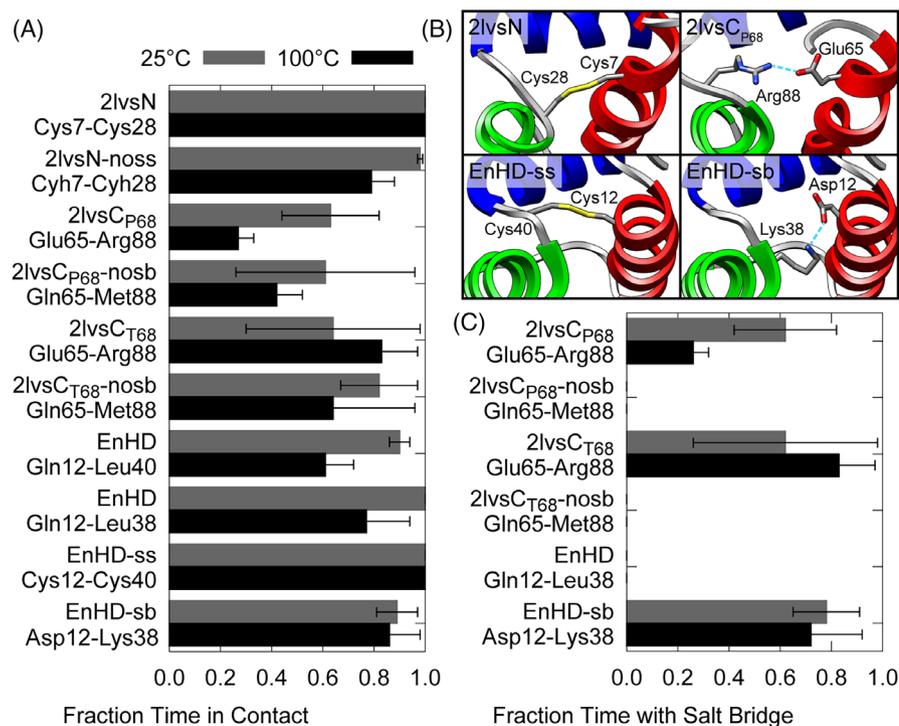
A common feature of thermostable proteins is an increase in surface charged residues, salt bridges, and hydrogen bonds compared to mesophilic proteins.<sup>10,13,40,43,44</sup> Indeed, 2lvsN had more surface hydrogen bonds than EnHD, but 2lvsC did not (Figure 4D). As the solvent density decreases at 100°C, hydrogen bonds with and among solvent become less common and additionally less favorable.<sup>24</sup> 2lvs lost surface hydrogen bonds with water at 100°C, but it gained them with itself, aiding in thermostabilization. The same was not true for EnHD, which lost hydrogen bonds at high temperature. EnHD and 2lvsC had the same number of unique surface contacts per residue, but 2lvsN had more (Figure 4C). An increased number of unique contacts indicates a more dynamic surface where residues can form more short-lived contacts with neighboring residues.

### 3.7 | The role of a buried salt bridge and disulfide bond

We aimed to investigate the role of 2lvsN's disulfide and 2lvsC's buried salt bridge by removing them from 2lvs and adding them to EnHD. In 2lvsN-noss, we simply reduced the disulfide bond and protonated both cysteine residues. In 2lvsC-nosb, we introduced two mutations, Glu65Gln and Arg88Met. Appropriate locations in EnHD between H1 and the very C-terminus of H11 were explored to find mutations that would geometrically support addition of a disulfide bond or salt bridge. We selected Gln12 and Leu40 to mutate to Cys in EnHD-ss, and Gln12 and Leu38 to mutate to Asp and Lys in EnHD-sb (Figure 5E).



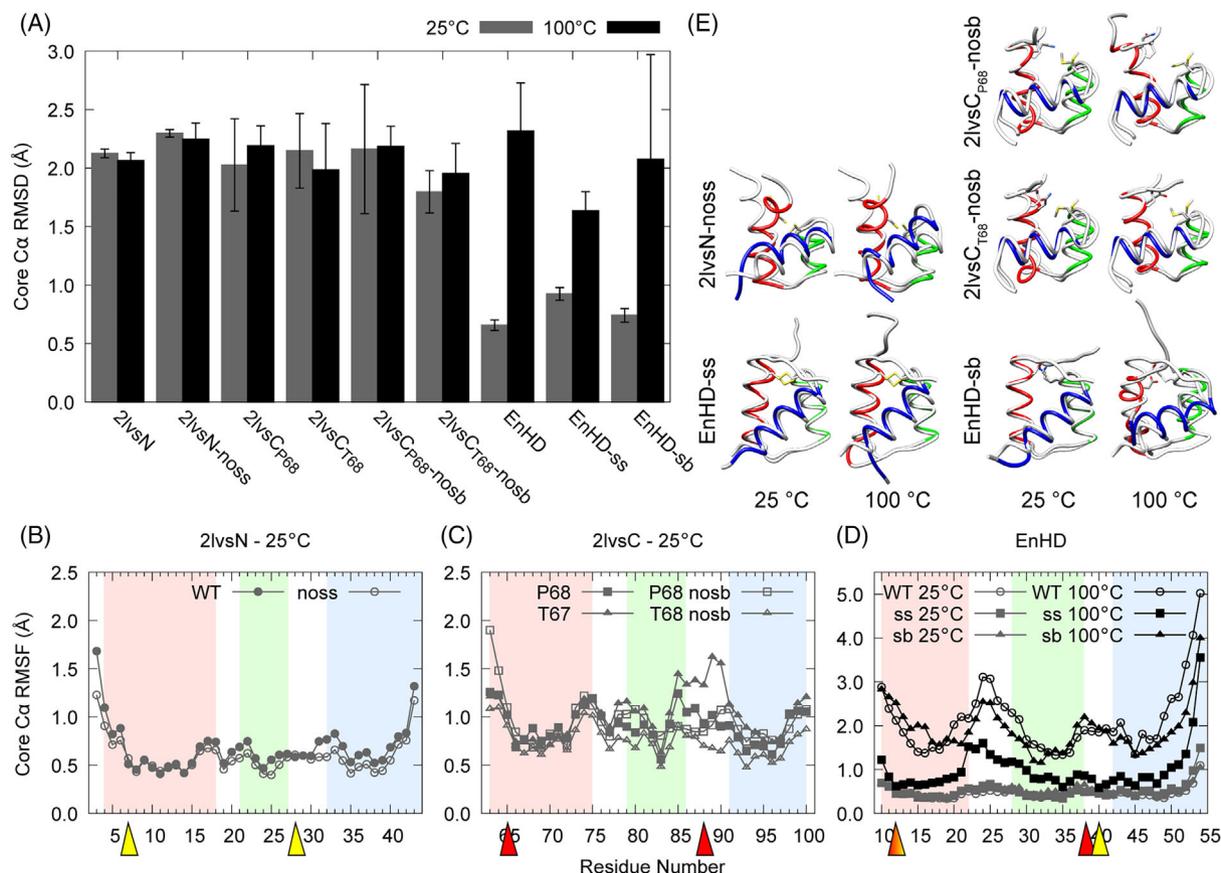
**FIGURE 4** Contribution to stability and quantification of side chain dynamics by residue-residue contacts. (A) Number of unique contacts among pairs of buried residues normalized by the number of buried residues in each protein. (B) Number of hydrogen bonds between pairs of buried residues. (C) Number of unique contacts among pairs of surface residues normalized by the number of surface residues. (D) Number of hydrogen bonds between pairs of surface residues. Average  $\pm$  SD,  $n = 5$



**FIGURE 5** Contacts between residues participating in native and engineered disulfide bonds and salt bridges. (A) Fraction time in contact for residues involved in the 2lvsN disulfide bond and 2lvsC salt bridge in the native and engineered proteins. (B) Images of the native disulfide bond (2lvsN) and salt bridge (2lvsCP68, dashed line) and those engineered into EnHD. (C) Fraction time with a salt bridge present for 2lvsC and salt bridge mutants. Average  $\pm$  SD,  $n = 5$

While the disulfide bond seemed to have provided some stability to 2lvsN, it was not critical for its global thermostability. Reduction of the disulfide bond in 2lvsN-noss increased somewhat the  $C_{\alpha}$  RMSD at 25°C but did not significantly change it at 100°C (Figure 6A), indicating the disulfide was not necessary for 2lvsN's thermostability. Reduction of the disulfide also did not decrease the amount of time the two residues spent in contact at 25°C. However, Cyh7 and Cyh28 in 2lvsN-noss spent only 79% of simulation time in contact at 100°C

(Figure 5A). We hypothesized that the disulfide bond might be responsible for side chain rigidification in 2lvsN's near-fully hydrophobic core, but we observed the same number of unique buried contacts/residue in 2lvsN-noss ( $2.2 \pm 0.1$  contacts/residue) as 2lvsN ( $2.2 \pm 0.0$  contacts/residue). Curiously, removal of the disulfide rigidified H3 somewhat (Figure 6B), perhaps providing further evidence that 2lvsN's backbone dynamics correlate with thermostability. While loss of 2lvsN's disulfide had little effect in our MD simulations at 100°C, in vitro mutation of



**FIGURE 6** Backbone dynamics of EnHD, 2lvs, and engineered mutants. (A) Average  $\pm$  SD C $\alpha$  RMSD over the core residues for all nine proteins at 25 and 100°C,  $n = 5$ . (B–D) Average C $\alpha$  RMSF over the core residues showing the effect of removal and addition of disulfide bonds and salt bridges. The location of the mutated Cys residues are noted with yellow arrows and mutated charged residues with red. Note the larger y-axis range in EnHD (D) compared to 2lvs (B, C). (E) Representative final MD structures (colored by helix) compared to the minimized starting structure (white)

the two cysteine residues to serine decreased the full protein's function above 50°C in DNA-binding experiments, suggesting the disulfide bond does contribute to 2lvsN's thermostability.<sup>18</sup>

The Glu65-Arg88 salt bridge in 2lvsC was unnecessary for the protein's stability, as the Glu65Gln and Arg88Met mutations did not significantly change the amount of time these residues spent in contact for either of the 2lvsC-nosb proteins at either temperature (Figure 5A). Similarly, there was no increase in core C $\alpha$  RMSD for either of the mutants (Figure 6A). If anything, 2lvsC<sub>T68</sub>-nosb was slightly more stable than the wild type. As with 2lvsN, changes in dynamics were most apparent in H6, with the salt bridge promoting rigidity in 2lvsC<sub>P68</sub> and flexibility in 2lvsC<sub>T68</sub> (Figure 6C).

Addition of a disulfide bond between residues 12 and 38 in EnHD was more stabilizing than addition of a salt bridge between residues 12 and 40. In EnHD, Gln12 and Leu40 were already in contact 90% of the time at 25°C but only 61% of the time at 100°C (Figure 5A). Addition of a disulfide bond between these residues increased both of these contact times to 100%, by definition. As for global stability, EnHD-ss was a bit more dynamic at 25°C based on core C $\alpha$  RMSD/F, but it was considerably more stable at 100°C (Figure 6A,D). Both EnHD-ss and EnHD-sb maintained helical content better at 100°C

than the wild type (Figure S3). Similarly, Gln12 and Leu38 were already in contact 100% of the time at 25°C and 77% of the time at 100°C. In EnHD-sb, this contact time decreased somewhat at 25°C but not significantly at 100°C (Figure 5A). As the measurement in Figure 5A was not specific to salt bridges, we wanted to ensure our salt bridge formed as designed. Measuring only contacts between the charged atoms in the amino acid pairs, the salt bridge was indeed formed 78 and 72% of the time at 25 and 100°C, respectively. Despite successful formation of the salt bridge, it did not obviously increase stability at 100°C based on core C $\alpha$  RMSD/F (Figure 6A,D,E).

While the engineered disulfide bond rigidified EnHD-ss at 100°C, stabilization was less obvious for EnHD-sb, based on simulation-wide averages. We hypothesized that perhaps the added interactions would slow unfolding at 100°C. We calculated transition state (TS) times for the three proteins in each simulation at 100°C. Wild type EnHD unfolded in all five of the simulations with TS times ranging from 25 to 87 ns (Table 1). The unfolding half-time for EnHD measured using laser T-jump experiments was extrapolated to 5 ns at 100°C, which is roughly an order of magnitude faster than we observed in our simulations.<sup>45</sup> However, our single-molecule simulations are not subject to the effects of protein-protein interactions, particularly those between

**TABLE 1** Unfolding transition state times in MD simulations at 100°C

Replicate	Transition state time (ns)		
	EnHD	EnHD-ss	EnHD-sb
1	87.42	71.08	>100 <sup>a</sup>
2	42.18	>100 <sup>a</sup>	85.15
3	37.42	91.92	62.19
4	25.32	98.33	>100 <sup>a</sup>
5	63.37	>100 <sup>a</sup>	>100 <sup>a</sup>

<sup>a</sup>The protein did not denature during the 100-ns simulation.

the hydrophobic regions exposed upon unfolding, which complicates the direct comparison to experimental data.

For EnHD-ss, two of the proteins did not denature in 100 ns, and the other three had TS times later than most of those observed for EnHD. Similarly, EnHD-sb did not denature in three of the simulations and denatured in the latter half of the simulation in both cases. Therefore, we predict that EnHD-ss and EnHD-sb have a slower  $k_u$  than wild-type EnHD that may impart kinetic stability.

### 3.8 | Comparison of stabilizations strategies with other thermophilic proteins

Naturally thermostable proteins tend to have more disulfide bonds and salt bridges than mesophilic proteins in static structures,<sup>10–12,43</sup> and thermostabilization of mesophilic and thermophilic proteins through addition of salt bridges or disulfide bonds has been previously accomplished.<sup>46–48</sup> Pordea and coworkers as well as Hazra and coworkers thermostabilized two carbonic anhydrases (CAs) by strategic addition of salt bridges<sup>46</sup> and disulfide bonds,<sup>47</sup> as we did for EnHD-ss. Pordea et al. also found that while two naturally thermophilic CAs were less dynamic than the mesophilic CA based on C $\alpha$  RMSF, another from thermophilic *S. azurea* (SazCA), was more dynamic. SazCA and 2lvs are two examples of naturally thermostable proteins whose thermostability could be achieved—or at least is not hindered by—increased backbone dynamics in the native state.

Proteins lose conformational entropy as they proceed from the highly disordered unfolded state to the more-organized native state, and this loss of entropy is energetically unfavorable. In this way, heightened native state dynamics provides thermodynamic stability through the loss of less entropy upon folding (smaller  $\Delta S$ ).<sup>16</sup> Indeed, Marqusee and coworkers calculated stability curves for RNase H\* from thermophilic *T. thermophilus* and mesophilic *E. coli* and found that the thermophilic protein was stabilized at higher temperatures than its mesophilic counterpart due to a smaller  $T\Delta S$ .<sup>49</sup> Similarly, Jaenicke and coworkers compared the cold shock protein from thermophilic *T. maritima* to that of mesophilic *B. subtilis* and found that both proteins had similar folding rates but that thermophilic protein had a slower unfolding rate constant, suggesting entropic factors are responsible for the thermostabilization of the thermophilic cold shock protein.<sup>15</sup>

Increased native state dynamics has also been associated with thermostability in engineered proteins, such as Precambrian  $\beta$ -lactamases designed by ancestral sequence reconstruction.<sup>50</sup> It is reasonable to predict that the older the ancestor, the more thermostable, due to the high but decreasing temperature of Earth over that time.<sup>51</sup> Sanchez-Ruiz and coworkers engineered de novo Kemp eliminase activity into the ancestral enzymes by adding an aspartate to a hydrophobic pocket, and the pattern held that the older the ancestor, the more eliminase activity. Structurally, the older and more active proteins were also more flexible, but perhaps counter intuitively, they had a more preorganized active site. Despite the heightened dynamics of the oldest protein, its active site was preorganized for transition-state stabilization, suggesting that thermostability by dynamics is compatible with function.

UVF is another example of an engineered, thermostable protein with a highly dynamic native state.<sup>7,8,42</sup> 2lvsC and UVF both had heightened backbone dynamics in the native state that were maintained without unfolding in MD simulations at 100°C. Unlike UVF, 2lvsC did not have heightened buried side-chain dynamics, so it is possible that entropic stabilization plays a smaller role in 2lvs' thermostability than UVF's. Like 2lvsC, UVF is a three-helix bundle, suggesting that stabilization by dynamics may not be specific to large enzymes and is perhaps a generalizable thermostabilization strategy.

Entropic thermostabilization through increased dynamics in the native state is a promising strategy for designing thermostable proteins. There are some examples of naturally thermostable proteins harnessing this form of thermostabilization, but the strategy is not straightforward to assess. Quantifying dynamics and thermodynamic contributions to stability require complicated and low-throughput techniques such as NMR, calorimetry, and molecular dynamics simulations. Additionally, increased dynamics are often assumed to indicate instability. More work determining the sequence and structural bases for thermostability via heightened nanosecond-timescale dynamics is needed for it to be a deployable tool for protein design.

## 4 | CONCLUSIONS

We investigated the structural and dynamic bases for thermostability in two thermophilic, three-helix-bundle, DNA-binding domains (2lvsN and 2lvsC) from *H. butylicus* and compared them to those previously observed in an engineered, thermostable three-helix bundle (UVF) and a naturally occurring, thermolabile three-helix bundle (EnHD). 2lvsN and 2lvsC had shorter and more-rigid loops between their first two helices compared to EnHD, and short loops are a common property of naturally thermophilic proteins. As is common in thermophilic proteins, 2lvsN had more hydrogen bonds on its surface than EnHD. Unlike other thermophilic proteins however, 2lvs buried a smaller fraction of its hydrophobic surface area upon folding and had a higher proportion of hydrophobic surface area compared to polar than EnHD. Strong core disulfide and salt-bridge interactions were not solely responsible for stability in 2lvs, but provided moderate (disulfide) or minimal (salt bridge) stability when engineered into a homologous position in EnHD. Like UVF, 2lvsC had heightened dynamics

relative to EnHD. We found that heightened nanosecond-timescale dynamics is compatible with thermostability, and indeed could be stabilizing. In addition to established methods of designing short, rigid loops, surface hydrogen bonds, and internal disulfide bonds, thermostabilization by dynamics may be a promising strategy for engineering thermostable proteins.

## ACKNOWLEDGMENTS

This work was supported by startup funds from the Santa Clara University College of Arts and Sciences, the DeNardo Research Foundation, and award number R15-GM134439 from the National Institutes of Health. The authors would like to thank Manasi Bhate for helpful comments on the manuscript.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Michelle E. McCully  <https://orcid.org/0000-0002-9774-0608>

## REFERENCES

- Basanta, M. J. Bick, A. K. Bera, C. Norn, C. M. Chow, L. P. Carter, I. Goreshnik, F. Dimairo, D. Baker, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 22135.
- Goldenzweig, S. J. Fleishman, *Annu. Rev. Biochem.* **2018**, *87*, 105.
- G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith, D. Baker, *Science* **2017**, *357*, 168.
- I. V. Korendovych, W. F. DeGrado, *Q. Rev. Biophys.* **2020**, *53*, E3.
- D. Baker, *Prot. Sci.* **2019**, *28*, 678.
- U. Mayor, C. M. Johnson, V. Daggett, A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 13518.
- P. S. Shah, G. K. Hom, S. A. Ross, J. K. Lassila, K. A. Crowhurst, S. L. Mayo, *J. Mol. Biol.* **2007**, *372*, 1.
- C. Nguyen, J. T. Young, G. G. Slade, R. J. Oliveira, M. E. McCully, *Biophys. J.* **2019**, *116*, 621.
- C.-H. Chan, T.-H. Yu, K.-B. Wong, *PLoS One* **2011**, *6*, e21624.
- R. Jaenicke, G. Böhm, *Curr. Opin. Struct. Biol.* **1998**, *8*, 738.
- S. Kumar, C.-J. Tsai, R. Nussinov, *Protein Eng., Des. Sel.* **2000**, *13*, 179.
- C. J. Reed, H. Lewis, E. Trejo, V. Winston, C. Evilia, *Archaea* **2013**, *2013*, 1.
- G. Vogt, S. Woell, P. Argos, *J. Mol. Biol.* **1997**, *269*, 631.
- I. N. Berezovsky, W. W. Chen, P. J. Choi, E. I. Shakhnovich, *PLoS Comput. Biol.* **2005**, *1*, e47.
- B. Schuler, W. Kremer, H. R. Kalbitzer, R. Jaenicke, *Biochemistry* **2002**, *41*, 11670.
- A. Karshikoff, L. Nilsson, R. Ladenstein, *FEBS J.* **2015**, *282*, 3899.
- W. Zillig, I. Holz, D. Janekovic, H. P. Klenk, E. Imsel, J. Trent, S. Wunderl, V. H. Forjaz, R. Coutinho, T. Ferreira, *J. Bacteriol.* **1990**, *172*, 3959.
- C. S. Kenchappa, P. O. Heidarsson, B. B. Kragelund, R. A. Garrett, F. M. Poulsen, *Nucleic Acids Res.* **2013**, *41*, 3424.
- N. D. Clarke, C. R. Kissinger, J. Desjarlais, G. L. Gilliland, C. O. Pabo, *Protein Sci.* **1994**, *3*, 1779.
- R. L. Dunbrack Jr., *Curr. Opin. Struct. Biol.* **2002**, *12*, 431.
- E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605.
- W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.
- H. Nagai, K. Kuwabara, G. Carta, *J. Chem. Eng. Data* **2008**, *53*, 619.
- R. Wolfenden, C. A. Lewis, Y. Yuan, C. W. Carter, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7484.
- J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
- A. D. MacKerell Jr., M. Feig, C. L. Brooks 3rd., *J. Comput. Chem.* **2004**, *25*, 1400.
- A. D. MacKerell Jr., D. Bashford, D. Bellott, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*, 3257.
- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1998**, *79*, 926.
- D. Beglov, B. Roux, *J. Chem. Phys.* **1998**, *100*, 9050.
- D. A. C. Beck, M. E. McCully, D. O. V. Alonso, V. Daggett, In *Lucem molecular mechanics (lmm)*, University of Washington, Seattle **2000-2021**.
- B. Lee, F. M. Richards, *J. Mol. Biol.* **1971**, *55*, 379.
- W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.
- A. Li, V. Daggett, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10430.
- A. Li, V. Daggett, *J. Mol. Biol.* **1996**, *257*, 412.
- R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria **2018**.
- M. E. McCully, D. A. C. Beck, V. Daggett, *Biochemistry* **2008**, *47*, 7079.
- N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, *Nature* **2012**, *491*, 222.
- N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S.-Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, W. F. DeGrado, *Nat. Chem.* **2017**, *9*, 1157.
- S. Kumar, R. Nussinov, *Cell. Mol. Life Sci.* **2001**, *58*, 1216.
- N. T. Southall, K. A. Dill, A. D. J. Haymet, *J. Phys. Chem. B* **2002**, *106*, 521.
- M. E. McCully, D. A. C. Beck, V. Daggett, *Prot. Eng. Des. Sel.* **2013**, *26*, 35.
- A. V. Glyakina, S. O. Garbuzynskiy, M. Y. Lobanov, O. V. Galzitskaya, *Bioinformatics* **2007**, *23*, 2231.
- T. B. Mamonova, A. V. Glyakina, O. V. Galzitskaya, M. G. Kurnikova, *BBA-Proteins Proteom.* **1834**, 2013, 854.
- U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. Freund, D. O. V. Alonso, V. Daggett, A. R. Fersht, *Nature* **2003**, *421*, 863.
- S. K. Bharatiy, M. Hazra, M. Paul, S. Mohapatra, D. Samantaray, R. C. Dubey, S. Sanyal, S. Datta, S. Hazra, *ACS Omega* **2016**, *1*, 1081.
- R. Parra-Cruz, C. M. Jäger, P. L. Lau, R. L. Gomes, A. Pordea, *J. Phys. Chem. B* **2018**, *122*, 8526.
- M. G. Pikkemaat, A. B. M. Linssen, H. J. C. Berendsen, D. B. Janssen, *Prot. Eng. Des. Sel.* **2002**, *15*, 185.
- J. Hollien, S. Marqusee, *Biochemistry* **1999**, *38*, 3831.
- V. A. Rizzo, S. Martinez-Rodriguez, A. M. Candel, D. M. Krüger, D. Pantoja-Uceda, M. Ortega-Muñoz, F. Santoyo-Gonzalez, E. A. Gaucher, S. C. L. Kamerlin, M. Bruix, J. A. Gavira, J. M. Sanchez-Ruiz, *Nat. Commun.* **2017**, *8*, 16113.
- V. Nguyen, C. Wilson, M. Hoemberger, J. B. Stiller, R. V. Agafonov, S. Kutter, J. English, D. L. Theobald, D. Kern, *Science* **2017**, *355*, 289.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** C. Nguyen, L. M. Yearwood, M. E. McCully, *J. Comput. Chem.* **2022**, *43*(3), 197. <https://doi.org/10.1002/jcc.26782>