A Dynamic Hydrophobic Core and Surface Salt Bridges Thermostabilize a Designed Three-Helix Bundle

Catrina Nguyen,[1] Jennifer T. Young,[1] Gabriel G. Slade,[2] Ronaldo J. Oliveira,[2] and Michelle E. McCully[1,*]

[1]Department of Biology, Santa Clara University, Santa Clara, California and [2]Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, Minas Gerais, Brazil

ABSTRACT Thermostable proteins are advantageous in industrial applications, as pharmaceuticals or biosensors, and as templates for directed evolution. As protein-design methodologies improve, bioengineers are able to design proteins to perform a desired function. Although many rationally designed proteins end up being thermostable, how to intentionally design de novo, thermostable proteins is less clear. UVF is a de novo-designed protein based on the backbone structure of the Engrailed home-odomain (EnHD) and is highly thermostable ($T_m > 99°C$ vs. $52°C$ for EnHD). Although most proteins generally have polar amino acids on their surfaces and hydrophobic amino acids buried in their cores, protein engineers followed this rule exactly when designing UVF. To investigate the contributions of the fully hydrophobic core versus the fully polar surface to UVF's thermostability, we built two hybrid, chimeric proteins combining the sets of buried and surface residues from UVF and EnHD. Here, we determined a structural, dynamic, and thermodynamic explanation for UVF's thermostability by performing 4 μs of all-atom, explicit-solvent molecular dynamics simulations at 25 and 100°C, Tanford-Kirkwood solvent accessibility Monte Carlo electrostatic calculations, and a thermodynamic analysis of 40 temperature runs by the weighted-histogram analysis method of heavy-atom, structure-based models of UVF, EnHD, and both chimeric proteins. Our models showed that UVF was highly dynamic because of its fully hydrophobic core, leading to a smaller loss of entropy upon folding. The charged residues on its surface made favorable electrostatic interactions that contributed enthalpically to its thermostability. In the chimeric proteins, both the hydrophobic core and charged surface independently imparted thermostability.

INTRODUCTION

Proteins in thermophilic organisms often have more salt bridges on their surface, helix-capping residues, and more-rigid loops relative to mesophilic proteins (1–4). Proteins constructed from ancestral or consensus sequences are generally more thermostable than their modern-day counterparts (5–7). Protein designers have taken advantage of these characteristics to intentionally thermostabilize otherwise-thermolabile proteins. Curiously, de novo-designed proteins often end up being thermostable without any direct effort on the designers' part to make them so (8–10). As protein-design methods improve, it will become possible to rationally design proteins that are not only functional but also thermostable. Here, we identify and investigate the contribution of two strategies to increase a protein's thermostability.

In the early 2000s, the Mayo Lab developed an algorithm to redesign the amino acid sequence of a protein based on a backbone template (11–13). As part of their algorithm, they classified all residue positions as "buried" or "surface" and allowed only hydrophobic residues to be placed at buried positions and only hydrophilic residues on the surface. As a proof of principle for their algorithm, they redesigned the sequence of the Engrailed homeodomain (EnHD), a three-helix bundle transcription factor from *Drosophila melanogaster*. The resulting protein, UVF, folded as expected and was unexpectedly, extremely thermostable ($T_m > 99°C$ vs. $52°C$ for EnHD) (14–17).

UVF and EnHD were the subject of a 2013 molecular dynamics study that investigated the structural and dynamic basis for thermostability in UVF (18). In those simulations, EnHD, as expected, is stable at 25°C and begins to unfold at 100°C. On the other hand, UVF is more dynamic than EnHD at room temperature and maintains these heightened dynamics at 100°C without unfolding. The authors predicted that the fully hydrophobic core of UVF contributes

Catrina Nguyen and Jennifer T. Young contributed equally to this work.

Editor: Alan Grossfield.

to its increased dynamics by allowing many favorable hydrophobic interactions, supporting more backbone conformations. Furthermore, they predicted that the more dynamic nature of UVF would be entropically favorable—UVF would lose less entropy when it folded relative to EnHD—contributing to UVF's thermostability.

Here, we test the prediction put forth in the 2013 study that UVF's hydrophobic core imparts an entropic thermostability by building two hybrid, chimeric proteins. Protein models of these chimeras were built by combining the sets of buried residues in UVF and surface residues in EnHD (UbEs) and vice versa (EbUs) (Fig. 1 *a*). We performed molecular dynamics simulations of all four proteins at 25 and 100°C to investigate their stability and dynamics. To test whether the hydrophobic core imparts an entropic stabilization, we performed coarse-grained simulations and calculated thermodynamic quantities (specific heat, enthalpy, entropy, free energy) using the weighted-histogram analysis method (WHAM) for the four proteins. UVF and both chimeras were more dynamic than EnHD at 25°C, and all three remained folded at 100°C. The observed thermostability of the chimeric proteins suggests that the hydrophobic core and hydrophilic surface of UVF both independently impart thermostability. Based on these results, as well as TKSA Monte Carlo electrostatic calculations, we propose that UVF's hydrophobic core provides entropic stabilization, and its hydrophilic surface provides enthalpic stabilization.

MATERIALS AND METHODS

Chimeric protein models

The chimeric protein models for EnHD buried UVF surface (EbUs) and UVF buried EnHD surface (UbEs) were built combining the buried (EnHD residue numbers 8, 12, 16, 19, 20, 26, 31, 34, 35, 38, 40, 44, 45, 48, 49, and 52) and surface (all other residues) residue sets of EnHD and UVF using MODELLER (Fig. 1) (19). The protein that provided the buried set of residues was used as the structure template in the homology model. An alignment file between the chimera and template was generated using align2d, and the automodel command was used to generate 6 (EbUs) or 50 (UbEs) models. The model with the lowest score by the MODELLER objective function was selected.

Molecular dynamics simulations

We performed all-atom, explicit-solvent molecular dynamics (AA-MD) simulations of EnHD, UVF, and both chimeric proteins. The starting structures for EnHD and UVF were the experimental crystal structure (Protein Data Bank (PDB): 1enh (20)) and NMR structure (PDB: 2p6j (14)), respectively. To assess the inherent stability of a crystal structure versus an NMR or model structure, we also performed simulations beginning from residues 3 to 56 of the NMR structure of EnHD (PDB: 2jwt (21)). All simulations were performed using NAMD 2.11 (22) with the CHARMM36m force field (23–25).

Structure files were generated for each of the five starting structures using autopsf in VMD (26), and 1000 steps of conjugate gradient minimization were performed. Each protein was solvated in a cubic box of TIP3P water molecules (27) of 50 Å/side (EnHD, UVF, chimeras) or 54 Å/side (EnHD NMR), and KCl was added at a concentration of 0.15 M to neutralize the system (28). The solvated system was minimized for an additional 100 steps and heated to 25 or 100°C. A Langevin thermostat and barostat were used to achieve an NPT system (constant number of particles, pressure, and temperature). The van der Waals forces were truncated with a smooth switching function at a cutoff distance of 8 Å, and full-system periodic electrostatics were performed using the particle mesh Ewald method. All bonds that included hydrogen were treated as rigid. An integration step size of 2 fs was applied, and structures were saved every 1 ps. Five 100-ns production runs were performed at both 25 and 100°C for EnHD (crystal), EnHD (NMR), UVF, EbUs, and UbEs for a total of 5 μs of simulation time.

Molecular dynamics simulation analysis

We performed data analysis using the molecular dynamics software, *in lucem* molecular mechanics (*il*mm) (29). All protein sequences were

a EnHD   UVF   EbUs   UbEs

b

c
```
           .    10    .    20    .    30    .    40    .    50    .
EnHD    RPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKI
UVF      MKQWSENVEEKLKEFVKRHQRITQEELHQYAQRLGLNEEAIRQFFEEFEQRK
EbUs     MKQFSENQEEKLKEEFKRHQRLTQEERHQLSQRLGLNEEQIRQWFEEKEQRK
UbEs    RPRTAWSSEVLARLKRFVNENRYITERRLQQYASELGLNEAAIKIFFQNFRAKI
                   [    HI    ]      [  HII  ]       [   HIII   ]
```
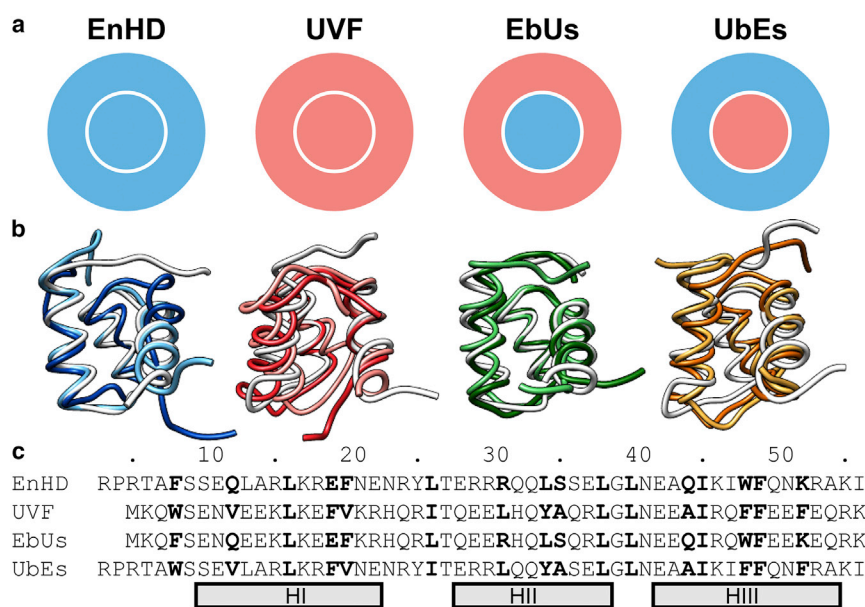
FIGURE 1 Structure and sequence of EnHD, UVF, and chimeric proteins. (*a*) Models depicting the design of EbUs and UbEs chimeras combining the surface and buried residue sets of EnHD and UVF. (*b*) Minimized starting structure (*white*) and representative final simulation structures at 25°C (*light*) and 100°C (*dark*) aligned on the helical core (residues 10–54). (*c*) Protein sequences with three helices (10–22, 28–38, 42–54) are noted below, and buried residues (8, 12, 16, 19, 20, 26, 31, 34, 35, 38, 40, 44, 45, 48, 49, 52) are noted in bold. To see this figure in color, go online.

numbered with respect to EnHD's residue numbering (Fig. 1 c). The root mean-square deviation (RMSD) of the Cα atoms relative to the minimized starting structure was calculated for the helical core (residues 10–54) of all proteins. These residues were selected to discount the flexibility of the disordered N- and C-termini and to allow direct comparison of Cα RMSDs between proteins. The core Cα RMSD was also calculated pairwise between structures across the simulation at 10-ps granularity, rather than comparing only to the starting structure, and reported as the all-versus-all core Cα RMSD.

Experimental NOEs for EnHD (MR block ID: 434304) (21) and UVF (MR block IDs: 449656 and 449657) (14) were downloaded from the Biological Magnetic Resonance Data Bank NMR Restraints Grid (30). 654 restraints were found for residues 3–56 of EnHD, and 1266 restraints were found for UVF. Distances were calculated as the $1/r^6$-weighted average distance between relevant pairs of hydrogen atoms, and an NOE was considered satisfied if the weighted distance was less than 5 Å or the r_{far} provided in the restraint file, whichever was longer. Distances were averaged across each simulation individually, over the five simulations at each temperature, and for both temperatures together. Long-order NOEs were defined as those between pairs of residues with a contact order of at least six.

Several different contact analyses were performed. Contact maps were calculated based on the amount of time each pair of residues had at least one pair of atoms in contact. Two atoms were considered in contact when pairs of carbon atoms were closer than 5.4 Å and when any other non-hydrogen atoms were within 4.6 Å. Residue pairs that were in contact in the minimized starting structure were defined as native contacts and plotted below the diagonal on contact maps, and non-native contacts were plotted above.

Contacts between side chains were further classified as hydrogen bonds, hydrophobic interactions, or other. Hydrogen bonds were defined as sets of three atoms in which the donor and acceptor atoms were <2.6 Å apart and the donor-hydrogen-acceptor angle was within 45° of linearity. Hydrophobic contacts were defined as pairs of carbon atoms that were bound to at least one hydrogen atom and were <5.4 Å apart. Other contacts were defined as all pairs of nonhydrogen atoms that were <4.6 Å apart. Side-chain salt bridges between ionizable residues were classified as hydrogen bonds or other contacts, as defined previously, that occurred between the carboxyl oxygen atoms of glutamic acid (none of the four proteins contains an aspartic acid) and the guanidinium, amine, or imidazole nitrogen and hydrogen atoms of arginine, lysine, or histidine. The total number of residue pairs that made any type of side-chain contact >10% of the simulation time were reported as "unique contacts." More unique contacts are indicative of a more dynamic protein.

Electrostatic free-energy calculations

Electrostatic free-energy calculations were performed using the Tanford-Kirkwood solvent accessibility (TKSA) model. In this model, the protein is treated as a spherical cavity with dielectric constant (ε_p) and radius (b) surrounded by an electrolyte solution modeled according to Debye-Hückel

theory (31). Shire et al. (32) modified this model by incorporating a solvent accessibility term for each ionizable residue that upgraded the TKSA model. Once the electrostatic energy between the titratable residues is determined, it is possible to calculate the free energy ($\Delta G_N(\chi)$) for the protein native state in a given protonation state χ (33,34). It is assumed in the method that the unfolded state does not contribute to electrostatic interactions (35). The protonation states χ were calculated via Metropolis Monte Carlo algorithm, resulting in the TKSA-MC model (36).

Thermodynamic analysis of coarse-grained models

Structure-based models (SBMs), also known as Gō-models, are widely employed to study the dynamics of protein ensembles, especially protein-folding mechanisms, because of the success of the energy landscape theory (37). SBMs of proteins can be simplified using different levels of coarse graining by including only Cα (Cα-SBM) or all heavy (non-hydrogen) atoms (HA-SBM) (38). The Hamiltonian that gives the protein interaction energy is based on the geometry of its native state, such that the potential energy surface reaches its minimum at this reference state (39). Simulations were performed following a similar protocol from the previous work (38) with GROMACS 4.6.7 (40). Input files for the simulations were prepared with SMOG2 version 2.1 beta with standard options (41). Thermodynamic analysis was performed over 40 temperature runs from lower (fully folded states) to higher temperatures (unfolded states), including the transition temperature from the folded to unfolded state (T_f), by the WHAM with an in-house Python script (42). T_f was defined as the temperature of the specific heat peak between the folded and unfolded states for each protein. The folded state was chosen to be 0.95 T_f, and the unfolded state was chosen to be 1.05 T_f to calculate ΔH, ΔS, and ΔF.

RESULTS

NOE satisfaction

To validate our AA-MD simulations, we compared them with experimentally determined nuclear Overhauser effect (NOE) distances. NOEs were obtained from the Biological Magnetic Resonance Data Bank, and $1/r^6$ weighted distances were calculated for all hydrogen atoms involved in NOE pairs. Table 1 shows the percentage of NOEs satisfied when the distances were averaged over each simulation, all five simulations at a given temperature, and all 10 simulations. EnHD satisfied 93% of long-range (contact order ≥6) and 94% of all experimental NOEs in simulations performed at room temperature. EnHD satisfied fewer long-range NOEs, only 91%, when the temperature was raised

TABLE 1 Fraction of NOEs Satisfied

Temperature (°C)	Run	EnHD		UVF	
		Long Range[a]	All NOEs	Long Range[a]	All NOEs
Minimized starting structure[b]		0.92	0.93	0.81	0.94
25	Average[c]	0.93 ± 0.01	0.94 ± 0.00	0.83 ± 0.03	0.94 ± 0.01
	All simulations	0.93	0.94	0.88	0.96
100	Average[c]	0.89 ± 0.03	0.94 ± 0.01	0.91 ± 0.02	0.96 ± 0.01
	All simulations	0.91	0.95	0.92	0.97
All simulations		0.95	0.96	0.91	0.97

[a]$i \rightarrow i + 6$ or greater.
[b]PDB: 1enh for EnHD, PDB: 2p6j for UVF.
[c]Average ± SD.

to 100°C because it began unfolding. UVF had lower long-range NOE satisfaction at 25°C than at 100°C, satisfying 88 vs. 92% long-range NOEs, respectively, because of faster sampling at the higher temperature. However, UVF satisfied 96 and 97% of all NOEs at 25 and 100°C, respectively. The number of long-range NOEs satisfied by the proteins in simulation was higher than the number satisfied by the starting structures, with 92% for EnHD and 81% for UVF, similarly because of increased sampling.

Backbone motion

The core Cα RMSD (residues 10–54) to the minimized starting structure was reported for a representative simulation for EnHD, UVF, EbUs, and UbEs and averaged across the five AA-MD simulations for each protein at each temperature (Fig. 1, *a* and *b*). RMSD data for all simulations are provided in Fig. S1. Higher RMSDs indicate more deviation from the starting structure, and increasing RMSD over time indicates that the protein is unstable and beginning to unfold. EnHD was the only protein that denatured at 100°C, on average reaching a core Cα RMSD of ~3.5 Å by 100 ns, compared with an average of 0.7 Å at 25°C. UVF and the two chimeras all had higher average core Cα RMSDs than EnHD at 25°C, but their RMSD did not increase much at 100°C. This backbone motion is evident in representative final structures from simulations at 25 and 100°C for all four proteins (Fig. 1 *b*). EnHD showed more movement from the starting structure (*white*) at 100°C (*dark blue*) than at 25°C (*light blue*), mostly in HIII, and the other three proteins show minimal differences between the two final structures.

Measuring RMSD to the starting structure to quantify stability assumes that the starting structure is not only accurate but represents a stable, native conformation. Because both starting structures for the chimeras were models, we wanted to ensure that the elevated RMSD we observed for the native state was not due to the proteins finding a different conformation and fluctuating stably in it. Therefore, we calculated the core Cα RMSD pairwise across all time points for all 10 simulations, an all-versus-all core Cα RMSD. These results are presented as heat maps and as averages of the five simulations at each temperature (Fig. 2, *c* and *d*). Larger, higher-resolution versions of these heat maps are provided in Fig. S1. At room temperature, EnHD had an average all-versus-all core Cα RMSD of 0.7 Å, whereas UVF, EbUs, and UbEs had more highly dynamic native states measuring 1.3, 1.3, and 1.7 Å RMSD, respectively. Although EnHD was much more dynamic at 100°C relative to 25°C, with a jump in RMSD of 2.2 Å, the other three proteins had modest increases in RMSD, all less than 0.8 Å. The all-versus-all RMSD heat maps additionally illustrate the variation in RMSD between temperatures. For UVF, EbUs, and UbEs, all the pairwise RMSDs between simulations at 25 and 100°C were not
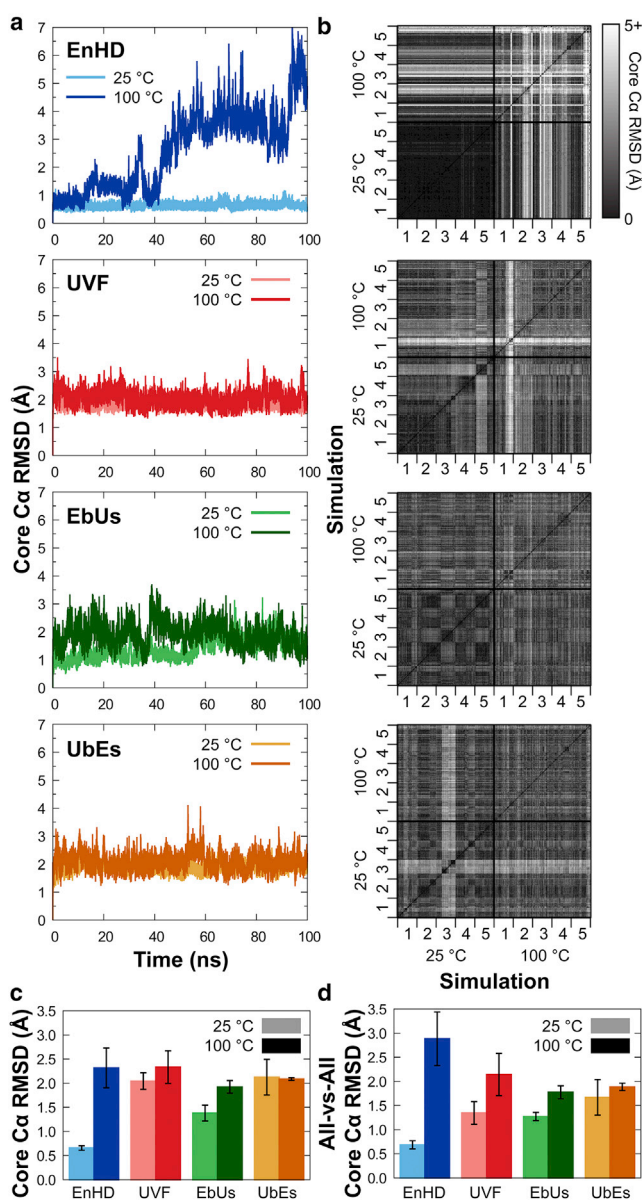
notably different than the pairwise RMSDs within simulations at the same temperature. Therefore, the conformational space sampled at 25 and 100°C was similar for UVF and the chimeras. EnHD, on the other hand, was very stable at 25°C, adopting fewer varied native conformations than the other three proteins, but it began exploring the denatured state at 100°C.

FIGURE 2 Core Cα RMSD. Cα RMSD of the helical core (residues 10–54) at 25 and 100°C is shown. (*a*) RMSD from one representative simulation of EnHD, UVF, EbUs, and UbEs at 25°C (*light*) and 100°C (*dark*). (*b*) All-versus-all core Cα RMSD matrices comparing structures at each time point pairwise across five simulations at both 25 and 100°C. (*c*) Average core Cα RMSD of the five simulations of at 25 and 100°C with error bars indicating SD, n = 5. (*d*) Average pairwise core Cα RMSD for all pairs of structures in each simulation, averaged over all runs, n = 5. For larger matrix images and RSMD versus time plots for all simulations, see Fig. S1. To see this figure in color, go online.

Native and non-native contacts

The average time two residues spent in contact over the five AA-MD simulations at both temperatures was plotted as a contact map to compare how dynamic each protein was (Fig. S2). The contacts were plotted above or below the diagonal based on whether they were present in the starting structure (native, below the *diagonal*) or gained during the simulation (non-native, above the *diagonal*). At 25°C, EnHD had few non-native contacts, with most occurring in the same regions of the map as the native contacts. This pattern indicates slight movements of the helices without large conformational changes. However, at 100°C, EnHD gained many non-native contacts that were smeared across the helices, and its native contacts had a lower percent time in contact compared to 25°C. Together, these observations indicate more drastic changes in conformation at 100°C than at 25°C, consistent with the early stages of unfolding.

At 25 and 100°C, UVF had more non-native contacts than EnHD did at 25°C. However, the non-native contacts were present in the same regions of the contact map as the native contacts, indicating that UVF made more unique residue-residue contacts than EnHD did at 25°C, whereas it maintained its folded structure at both temperatures. The contact patterns observed for EbUs and UbEs were most similar to those of UVF, suggesting that all three proteins were highly dynamic at both temperatures, with residues coming into contact with the residues next to those they made contacts with in the starting structures.

Side-chain contacts

To investigate what types of interactions were giving rise to the heightened dynamics observed in UVF and the two chimeric proteins in the AA-MD simulations, we investigated what types of interactions were made between side chains in different regions of the protein. Side-chain contacts were analyzed by type of contact (hydrogen bond, hydrophobic interaction, or other) and by the region to which both of the residues in the contact belonged (buried, surface, or buried ↔ surface) (Fig. 3).

EnHD lost more total side-chain contacts than UVF, EbUs, and UbEs at 100°C relative to 25°C (Fig. 3 *p*). For EnHD, 66 of the 117 total contacts lost at 100°C relative to 25°C were buried hydrophobic interactions (Fig. 3, *b* and *p*), which is consistent with the early events of unfolding. The other three proteins lost closer to 30 hydrophobic interactions when the temperature was raised and only lost 11–46 contacts overall.

UVF and EbUs both had more surface hydrogen bonds than EnHD and UbEs (Fig. 3 *e*), and both gained an average of about one hydrogen bond when the temperature was raised. On the other hand, EnHD and EbUs both had three to four hydrogen bonds between surface and buried residues, whereas UVF and EbUs had none (Fig. 3 *i*). UVF and UbEs had fewer buried "other" contacts than EnHD or EbUs (Fig. 3 *c*). This type of contact includes interactions between charged groups that do not have the proper geometry of a hydrogen bond as well as interactions between polar and hydrophobic groups.

The number of unique contacts that a protein makes is related to how dynamic it is. We totaled the number of unique buried and surface residue pairs whose side chains were in contact for during at least 10% of the simulation time to quantify the dynamics of the hydrophobic core and hydrophilic surface in each of the four proteins (Fig. 3, *q* and *r*). EnHD was the most rigid, with 49 unique buried contacts at 25°C, whereas the other proteins had 53–59. Again, the number of unique buried contacts increased by 13 at 100°C relative to 25°C, as EnHD began denaturing, whereas there was an increase of only one to four contacts for UVF and the two chimeric proteins. EnHD and UbEs both have two more surface residues than UVF and EbUs (Fig. 1 *c*), in part accounting for their higher number of unique surface contacts. However, both chimeras have five to six more unique surface contacts at 25°C than the protein from which their surface was derived.

EnHD crystal versus NMR structure

EnHD was more stable than UVF and both chimeras, but its simulation starting structure was a crystal structure, whereas UVF's was an NMR structure, and the two chimeras' structures were homology models. To assess the contribution of structure determination method to protein stability as observed in our simulations, we compared simulations of EnHD beginning from the crystal (PDB: 1enh) or NMR (PDB: 2jwt) structure. 2jwt was as stable as 1enh based on core Cα RMSD to the starting structure as well as pairwise across simulations (Figs. 2 and S3, *a–c*). 2jwt and 1enh satisfied the same number of NOEs at both temperatures (Tables 1 and S1). Like 1enh, 2jwt began making non-native contacts in regions outside of the three-helix bundle topology in simulations at 100°C but maintained contacts within the topology at 25°C (Figs. S2 and S3 *d*). 2jwt made slightly fewer hydrogen bonds and "other" contacts than 1enh at 25°C, but their contacts were identical at 100°C (Figs. 3 and S3 *e*). 2jwt made more unique buried contacts at 25°C than 1enh, but not as many as UVF, and the same number of unique surface contacts (Figs. 3, *r* and *q* and S3 *f*).

Electrostatic interactions

EnHD and UVF have about the same number of charged residues (Arg, Lys, Glu; 19 and 21, respectively; neither protein has an Asp), but UVF has a more equal balance between positive and negative residues (EnHD net charge +7, UVF −1, Fig. 1 *c*). We investigated how this change in charge
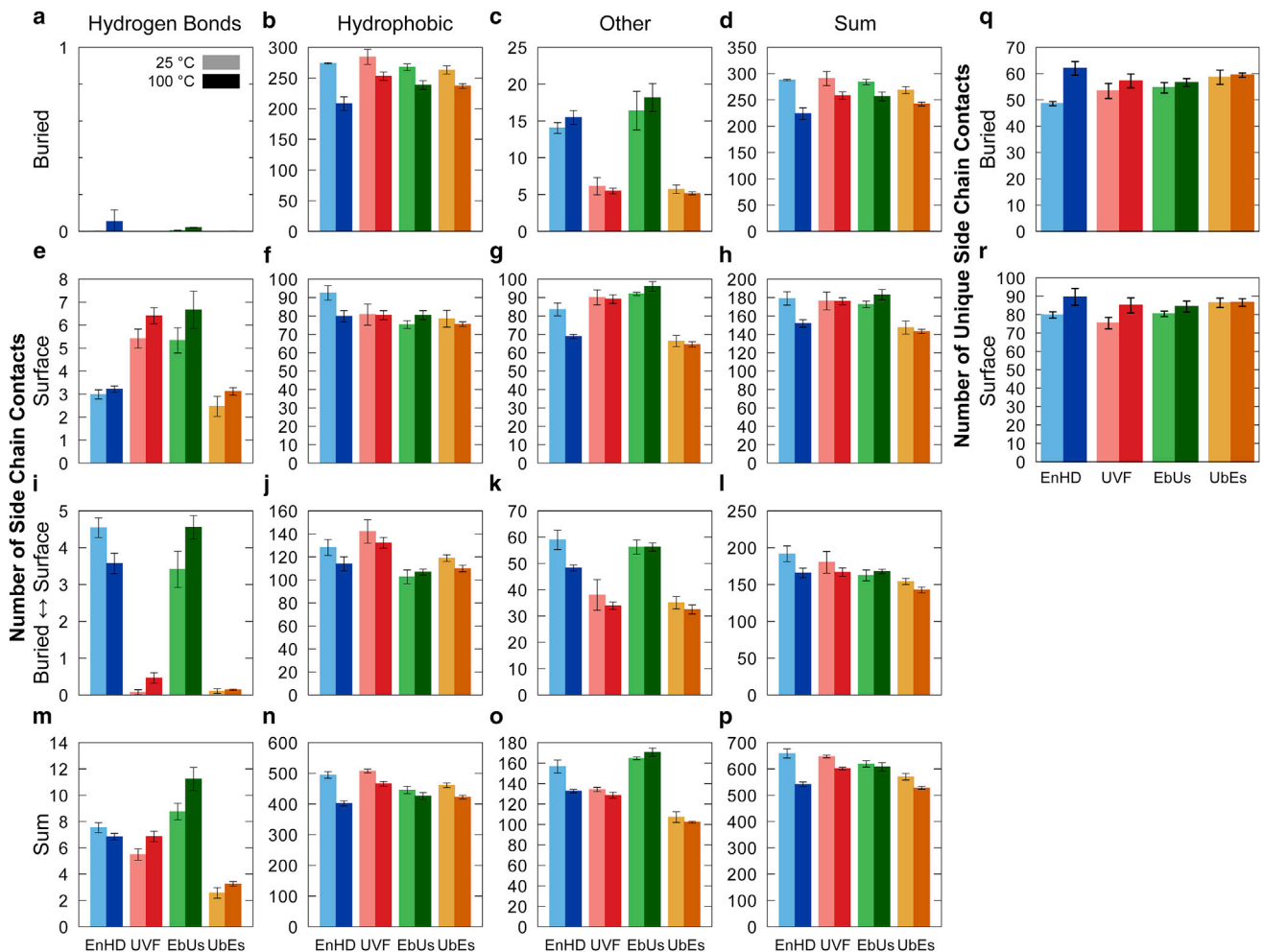
FIGURE 3 Side-chain contacts by region and contact type. Side-chain contacts are categorized as between residues that are (*a–d*) both buried (residues: 8, 12, 16, 19, 20, 26, 31, 34, 35, 38, 40, 44, 45, 48, 49, 52), (*e–h*) both surface (residues: 3–7,9-11, 13–15, 17, 18, 21–25, 27–30, 32, 33, 36, 37, 39, 41–43, 46, 47, 50, 51, 53–56), or (*i–l*) one buried and one surface. Total side-chain contacts (*m–p*) are also displayed. Contacts are further typed as (from *left* to *right*) hydrogen bonds, hydrophobic contacts, other interactions, and the sum of those three groups. Unique residue side-chain pairs in contact >10% of the simulation time are plotted for (*q*) buried and (*r*) surface residues as well. Number of contacts was averaged across five simulations at 25°C (*light*) and 100°C (*dark*), and error bars indicate SD. To see this figure in color, go online.

distribution affected the contacts that could be formed as well as the energetic contribution of electrostatic interactions.

All five Glu residues in EnHD formed a salt bridge with an Arg residue 100% of the time in AA-MD at 25°C (Table 2). UVF, on the other hand, had only two contacts present 100% of the simulation time at 25°C. UVF made eight salt bridges during ≥80% of the AA-MD simulation time, involving 12 of 23 distinct ionizable residues, whereas EnHD made five salt bridges involving 7 of 19 ionizable residues, two of which were in the buried position. For the residues making salt bridges, they were equally efficient in UVF and EnHD (0.67 vs. 0.71 salt bridges/residue, respectively), but UVF involved a higher proportion of its ionizable residues in salt bridges than EnHD (52 vs. 37%).

The EbUs chimera made 13 salt bridges ≥80% of the simulation time at 25°C, but only one 100% of the time. Five of these salt bridges were between residue pairs that

also made salt bridges ≥80% of the time in UVF. EbUs was as efficient as UVF and EnHD at using residues for multiple salt bridges (13 salt bridges among 19 ionizable residues, 0.71 salt bridges/residue), but it had 19 of 22 ionizable residues making salt bridges (86%). UbEs, on the other hand, formed only two salt bridges ≥80% of the simulation time, both of which were observed <80% of the time in EnHD.

The TKSA-MC analysis allows the calculation of the charge-charge interaction free energy (ΔG_{qq}) of each ionizable residue. A residue with a positive ΔG_{qq} tends to destabilize the protein native state, whereas a negative ΔG_{qq} indicates that the residue is favorable to stabilize the protein. Fig. 4 shows TKSA-MC analysis for the EnHD, UVF, and the two chimeric proteins. In the EnHD, almost 50% of the ionizable residues were not favorable to the protein stability. However, UVF had a greater number of ionizable

TABLE 2 Side-Chain Salt Bridges between Ionizable Residues

Contact Pair	Fraction Time in Contact[a]	
	EnHD	UbEs
Glu28-Arg31Leu[b]	1.00	0.00
Glu19Phe-Arg30	1.00	0.00
Arg31Leu-Glu42	1.00	0.00
Arg15-Glu19Phe	1.00	0.00
Arg15-Glu37	1.00	0.06
Arg18-Glu22	0.64	0.90
Glu28-Arg53	0.00	0.95
	UVF	EbUs
Glu51-Arg55	0.96	0.90
His32-Glu42	0.92	1.00
Glu14-Lys17	0.91	0.99
Lys15-Glu18	0.88	0.92
Glu18-Arg22	0.86	0.87
Lys6-Glu13	1.00	0.59
Lys6-Glu10	1.00	0.28
Glu13-Lys17	0.80	0.57
Glu18-Lys21	0.73	0.89
Glu43-Arg46	0.57	0.81
Arg46-Glu50	0.47	0.88
His23-Glu30	0.18	0.93
Arg25-Glu53	0.08	0.98
Phe19Glu-His23	0.00	0.96
Phe19Glu-Arg37	0.00	0.97
Phe19Glu-Arg22	0.00	0.99

[a]For contacts present at least 80% of the time in at least one of the two proteins in five simulations at 25°C.
[b]Mutation notation gives EnHD or UVF residue first and chimera residue second.

residues, and almost all of them were electrostatically favorable. This analysis corroborates with the increase of the thermostability of UVF when compared with EnHD. Regarding the two chimeras, the free energies of UbEs's ionizable residues were similar to those of EnHD's, and EbUs's ΔG_{qq} distribution was more similar to UVF. The sum of each residue's ΔG_{qq} contribution is shown in Fig. 4 e for all four proteins, and the total ΔG_{qq} was lowest for UVF, followed by EbUs, EnHD, and UbEs.

Thermodynamic analysis

SBMs define the energy minimum of the Hamiltonian function as the protein's native conformation; thus, it is a topology-based model with fast sampling over a span of temperatures. SBM simulations were performed for the four proteins using two levels of coarse graining, Cα-SBM or HA-SBM, to quantify the thermodynamics of the folding process (Figs. 5 and 6). Fig. 5 compares the specific heat as a function of the simulated temperature range ($c_v(T)$) for the four proteins in the Cα-SBM (Fig. 5 a) and HA-SBM (Fig. 5 b). Only the HA-SBM simulations were able to show differences in the folding or melting temperature (T_f) of the studied proteins, given by the temperature of the peak in the $c_v(T)$ curve. This discrepancy is a conse-

quence of the side-chain interactions included in the HA-SBM, which are not included in the Cα-SBM. UVF and UbEs had a higher T_f than EnHD and EbUs (Fig. 5 b).

Fig. 6 shows the enthalpy (H, Fig. 6 a), entropy (S, Fig. 6 b), and free-energy (F, Fig. 6 c) profiles as a function of temperature obtained by the heavy-atom SBM simulations. As in Fig. 5 b, the higher UVF and UbEs enthalpy and entropy profiles culminate in the lower free-energy profile ($F = H − TS$) over the simulated temperature range (Fig. 6 c). These profiles were used to calculate the relative changes in enthalpy and entropy upon folding (Table 3). UVF and UbEs both lost less entropy than EnHD when they folded (smaller ΔS), and UVF and EbUs both gained more enthalpy than EnHD when they folded (larger ΔH). EnHD was the least stable with the smallest ΔF.

DISCUSSION

Consistent with experimental results, EnHD was stable in AA-MD simulations at 25°C and began unfolding at 100°C, whereas UVF was stable at both temperatures (Fig. 2). In coarse-grained SBMs, side-chain interactions, which were only included in the HA-SBM, were critical to reproducing the differences in melting temperatures observed for EnHD and UVF experimentally and in the AA-MD simulations (Fig. 5 b). The Cα-SBM is purely native-topology based, and because both proteins fold to nearly identical topologies, the level of theory was not detailed enough to reproduce relative differences in melting temperatures.

UVF has an entirely hydrophobic core and entirely polar surface. When we spliced together the cores and surfaces of UVF and EnHD (Fig. 1 a), both resulting chimeric proteins were stable over 100 ns of AA-MD simulation (Figs. 2 and S2). The melting temperatures derived from the HA-SBM for both chimeras were between those of EnHD and UVF, with EbUs more similar to EnHD and UbEs more similar to UVF (Fig. 5 b), and both chimeras had a larger ΔF than EnHD (Table 3). UbEs had heightened thermostability at high temperature in both the AA-MD and HA-SBM simulations. However, for EbUs, the HA-SBM gave a melting temperature only slightly higher than EnHD's, whereas the events of early unfolding were not observed by AA-MD at 100°C. It is possible that EbUs's folding kinetics are not fast enough for unfolding to be observed on the 100-ns timescale, though those of its parent proteins are ($t_{1/2} = 15$ μs for EnHD and 9 μs for UVF (15,17)).

Entropic stabilization by a dynamic hydrophobic core

The fully hydrophobic core of UVF contributed to the heightened stability of UbEs. Like UVF, UbEs was more dynamic than EnHD at room temperature and maintained
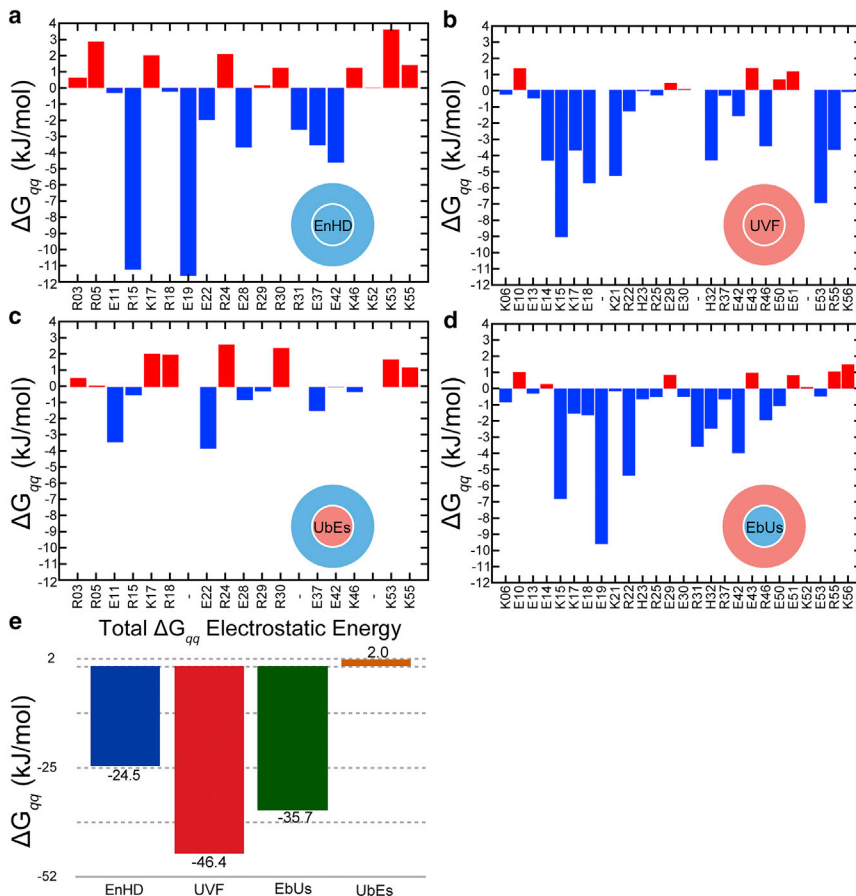
FIGURE 4 Electrostatic free energy. (a–d) Charge-charge interaction energy (ΔG_{qq}) calculated by the TKSA-MC model for each ionizable residue. Negative values indicate that the residue is electrostatically favorable, whereas positive values indicate that the residue is unfavorable to the electrostatic free energy. (e) Total ΔG_{qq} energy for each protein. To see this figure in color, go online.

these dynamics at 100°C, as evident in the higher core Cα RMSD relative to the starting structure (Fig. 2, a and c). The conformational spaces sampled by UVF and UbEs were very similar at both temperatures, as quantified by all-versus-all core Cα RMSD and contact maps (Figs. 2 b and S2). The heightened backbone dynamics could be tolerated without unfolding because of the molten, fully hydrophobic core, as measured by increased unique buried side-chain contact pairs relative to EnHD (Fig. 3 q). Replacing EnHD's buried polar residues with hydrophobic

ones in UVF and UbEs did not decrease the number of buried hydrophobic contacts, but it did remove eight unfavorable, buried "other" contacts of the 14 in EnHD (Fig. 3, b and c). Creating tightly packed hydrophobic cores is a known method of stabilizing proteins (43), but although the cores of UVF and UbEs are fully hydrophobic, the packing was loose and constantly rearranged.

UbEs had the fewest surface hydrogen bonds and the fewest hydrogen bonds overall among all four proteins. Curiously, addition of surface hydrogen bonds is known
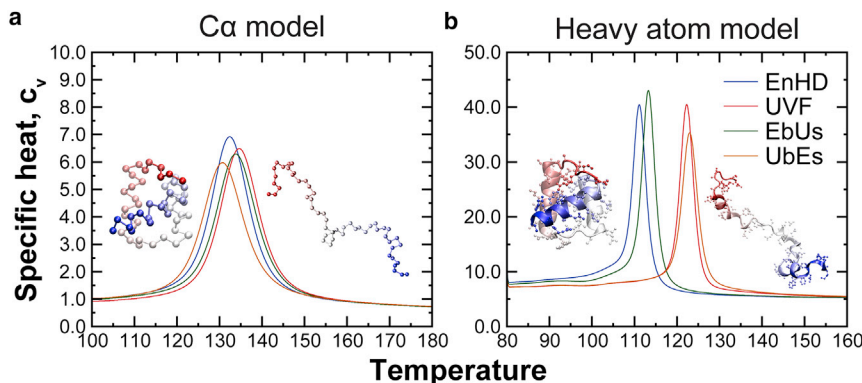

FIGURE 5 Specific heat profiles. Specific heat as a function of temperature ($c_v(T)$) for the SBM simulations in (a) Cα-SBM and (b) HA-SBM is shown. Structures of the folded native state (left) and one unfolded conformation (right) are shown for each model in (a) and (b). Energy is in units of $k_B T$, and temperature is in GROMACS reduced units. To see this figure in color, go online.
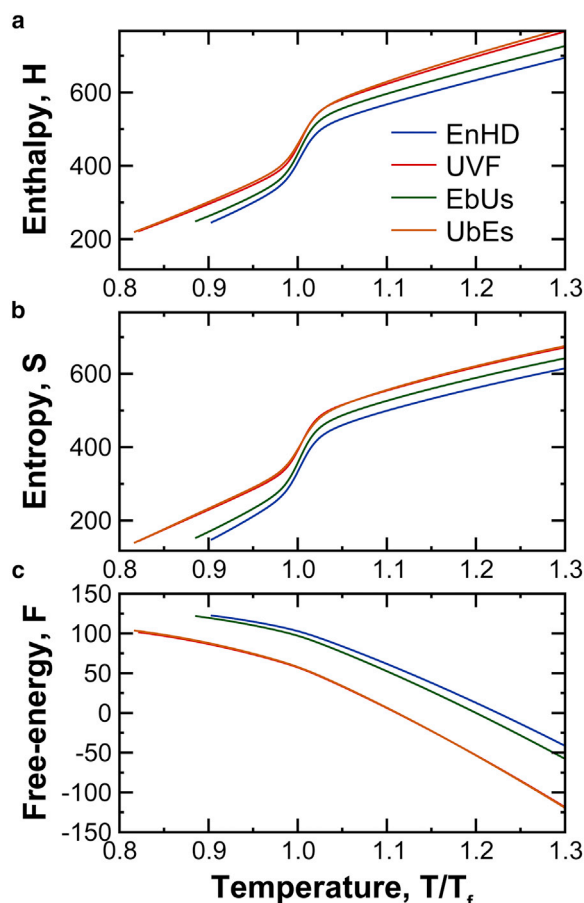
a

b

c

FIGURE 6 Enthalpy, entropy, and free-energy profiles. (*a*) Enthalpy ($H(T)$), (*b*) entropy ($S(T)$), and (*c*) free-energy ($F(T)$) profiles as a function of temperature for the HA-SBM simulations. Free energy is given by the thermodynamic relationship, $F = H - TS$. Energy is in units of k_BT, and temperature is rescaled by the folding temperature (T_f) of each protein. To see this figure in color, go online.

to stabilize proteins (4,43,44), yet UbEs achieved heightened thermostability relative to EnHD with one fewer surface hydrogen bond and five fewer total side-chain hydrogen bonds (Fig. 3, *e* and *m*). UbEs had a highly dynamic surface, similar to what was observed for its core, resulting in five more unique surface contacts than UVF and as many as EnHD despite having two fewer surface residues (Fig. 3 *r*). EnHD had five salt bridges present 100% of the simulation time (Table 2). UbEs had only two pre-

TABLE 3 Relative Thermodynamic Parameters for Folding

	ΔH	ΔS	ΔF
EnHD	229	248	−31
UVF	231	229	−40
EbUs	238	254	−34
UbEs	229	225	−41

Units are k_BT; folded and unfolded states from WHAM were defined as 0.95 and 1.05 T_f, respectively.

sent $\geq 80\%$ of the simulation time, one of which was not observed in EnHD.

The stabilization of UbEs was not due to increased salt bridges, on the surface or otherwise. The molten, fully hydrophobic core allowed both UVF and UbEs to be more dynamic at 25°C and maintain these dynamics without unfolding at 100°C. These heightened dynamics are observed in the HA-SBM thermodynamic analysis. The temperature-dependent enthalpy and entropy profiles of UbEs are similar to those of UVF and higher than those of EnHD (Fig. 6, *a* and *b*). Higher enthalpy indicates the system accesses higher energy states due to heat being absorbed, and entropy is a measure of the unavailability of a system's energy to do work. Thus, the absorbed heat of UVF and UbEs increased the proteins' entropy, which is reflected in the increased sampling of conformational space observed in the AA-MD simulations. UVF and UbEs were stabilized entropically by their highly heterogeneous native states. UVF and UbEs lost less entropy upon folding ($\Delta S = 229$ and 225 k_BT, respectively; Table 3) than EnHD and EbUs ($\Delta S = 248$ and 255 k_BT).

Enthalpic stabilization by surface electrostatic interactions

The fully polar surface of UVF contributed to the heightened stability of EbUs. UVF and EbUs both had more surface hydrogen bonds and overall more contacts on the surface than EnHD (Fig. 3, *e* and *h*). UbEs's surface contacts were especially dynamic; it had the most unique surface contact pairs as well as more and shorter-lived side-chain salt bridges (Fig. 3 *r*; Table 2).

Both EnHD and EbUs have hydrogen bonds between buried and surface residues as well as overall more "other" contacts than UVF (Fig. 3, *i* and *o*). We predicted that these interactions could be destabilizing to EnHD; however, EbUs was more stable than EnHD, so they could not have been too unfavorable. EbUs was also more dynamic that EnHD, with higher core and all-versus-all Cα RMSDs and more unique buried contact pairs (Figs. 2 and 3 *q*). These heightened dynamics could not have been due to the fully hydrophobic core as in UVF and UbEs, yet they seem to be similarly stabilizing.

The TKSA-MC model results showed similar electrostatic behavior for UVF and EbUs (Fig. 4, *b* and *d*), for which nearly all ionizable residues contribute favorably to their stability. The total ΔG_{qq} for these two proteins was more favorable than for EnHD and UbEs (Fig. 4 *e*). EnHD and UbEs both contained several ionizable residues that were not favorable to their stability, including Lys17, Arg24, Arg30, Lys53, and Lys55 (Fig. 4, *a* and *c*). This fact strengthens the argument that UVF's polar surface contributed to the stability of EbUs. The majority of ionizable residues in UVF and EbUs were involved in salt bridges in AA-MD simulations, unlike in EnHD (52, 86, and 37%,

respectively) (Table 2). UVF and EbUs also had overall more salt bridges (Table 2) and more stability imparted by electrostatic interactions (Fig. 4) than EnHD. UVF and EbUs gained stability from using their ionizable residues more efficiently, making more salt bridges than EnHD and including more ionizable residues in salt bridges.

The thermodynamic quantities calculated for EbUs with the HA-SBM were between those of UVF and EnHD (Fig. 6), although EbUs presented heightened stability in the AA-MD simulations (Fig. 1) and low charge-charge electrostatic free energy calculated by the TKSA-MC model (Fig. 4). UVF and EbUs had a greater change in enthalpy upon folding ($\Delta H = 231$ and 238 k_BT, respectively) than EnHD or UbEs ($\Delta H = 229$ k_BT for both), which may be due to their increased number of salt bridges and fraction of ionizable residues with satisfied electrostatic interactions (Table 3).

The HA-SBM simulations employed here do not account for electrostatics; they are a purely topology-based Hamiltonian. As discussed previously, we observed that side-chain salt bridges and electrostatic interactions contributed to EbUs's thermostability. Thus, topology-based side-chain interactions seem to play an important role in protein stability at high temperatures for UVF and UbEs, whereas EbUs was more affected by electrostatics than topology-based contacts interactions. It is also possible that EbUs is less stable than UVF and UbEs, which would indicate that the fully hydrophobic core is more stabilizing than the electrostatic surface.

Implications for protein engineering

Protein engineers typically design proteins by placing hydrophobic residues at buried positions and polar residues on the surface of proteins. In protein-design algorithms, tightly packed hydrophobic cores are rewarded in scoring functions (12,45). Proteins evolve naturally following the same, general scheme, but it is unusual, both in nature and in protein design, to end up with proteins that have exclusively hydrophobic residues at buried positions and polar residues on the surface, as is the case with UVF (46). Our AA-MD simulations showed that the fully hydrophobic core in both UVF and UbEs resulted in a more dynamic protein at 25°C that maintained the heightened dynamics without unfolding at 100°C. The dynamics of core residues observed previously in AA-MD simulations for EnHD and other proteins agree quantitatively with order parameters measured by NMR (18,47).

Similarly, hydrogen bonds and salt bridges are intentionally placed on the surface of proteins by rational protein designers, and scoring functions reward such interactions (13,45). Proteins that evolve naturally in thermophilic organisms tend to have more surface salt bridges than those in mesophiles (4,44). Salt bridges in UVF and EbUs were highly dynamic and more numerous compared to those in

EnHD; indeed, few were present for the full simulation time. Designed proteins typically are more stable than naturally evolved proteins (8,14,48). It remains to be seen whether designed proteins are typically more dynamic or whether there is a relationship between heightened dynamics and stability.

Thermodynamically, a highly dynamic, hydrophobic core that leads to a more heterogeneous native state should manifest as a smaller loss of entropy upon folding. The role of entropy in stabilizing thermophilic proteins has been investigated previously. For example, RNase H* from *Thermus thermophilus* is stabilized at higher temperatures than its *Escherichia coli* counterpart because of a smaller $T\Delta S$, likely resulting from a reduced ΔC_p and smaller hydrophobic effect (49). The cold shock protein from *Thermotoga maritima* unfolds slower than the *Bacillus subtilis* cold shock protein, indicating that it is stabilized entropically, most likely because of increased degrees of freedom in the native state (50).

Designing proteins with highly dynamic, hydrophobic cores seems consistently to produce thermostable proteins. Thermostable proteins are popular starting points for directed evolution and designing proteins to perform a desired function. Thermostable proteins can withstand mutations that are required for adopting a new function but may be destabilizing to the overall structure. It remains to be seen whether highly dynamic proteins such as UVF can serve as modifiable scaffolds. It is possible that this stabilization strategy is not compatible with function, especially if the protein might need to adopt a specific conformation to carry out its function.

CONCLUSIONS

Here, we determine a structural, dynamic, and thermodynamic explanation for the thermostability of a de novo-designed variant of EnHD using AA-MD simulations, TKSA Monte Carlo electrostatic calculations, and heavy-atom SBMs. Our models suggest that UVF is more dynamic because of its fully hydrophobic core, which results in a smaller loss of entropy upon folding. It is further stabilized by its favorable surface electrostatics, resulting in a larger gain of enthalpy upon folding. Based on our models of chimeric proteins that combine the cores and surfaces of EnHD and UVF, both the dynamics of UVF's fully hydrophobic core and the increased electrostatic interactions in its fully polar surface independently impart thermostability to the chimeras. All three computational models were necessary to identify and quantify the contribution of different structural patterns to the thermostability of the studied proteins, and together they provide, to our knowledge, a novel workflow for dissecting the contributions to stability in other thermostable proteins. Our results suggest a mechanism of protein thermostabilization that has not been thoroughly explored in the protein-design community,

but it remains to be seen whether it is compatible with protein function.

SUPPORTING MATERIAL

Three figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(19)30028-1.

AUTHOR CONTRIBUTIONS

M.E.M., C.N., and J.T.Y. designed the chimeric proteins, and they designed and conducted the molecular dynamics simulations. R.J.O. and G.G.S. designed and conducted the electrostatics calculations and structure-based model simulations. All authors analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

1. Russell, R. J., and G. L. Taylor. 1995. Engineering thermostability: lessons from thermophilic proteins. *Curr. Opin. Biotechnol.* 6:370–374.

2. Tokuriki, N., and D. S. Tawfik. 2009. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19:596–604.

3. Wijma, H. J., R. J. Floor, and D. B. Janssen. 2013. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* 23:588–594.

4. Jaenicke, R., and G. Böhm. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* 8:738–748.

5. Socha, R. D., and N. Tokuriki. 2013. Modulating protein stability - directed evolution strategies for improved protein function. *FEBS J.* 280:5582–5595.

6. Lehmann, M., L. Pasamontes, …, M. Wyss. 2000. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta.* 1543:408–415.

7. Tripp, K. W., M. Sternke, …, D. Barrick. 2017. Creating a homeodomain with high stability and DNA binding affinity by sequence averaging. *J. Am. Chem. Soc.* 139:5051–5060.

8. Rocklin, G. J., T. M. Chidyausiku, …, D. Baker. 2017. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science.* 357:168–175.

9. Chevalier, A., D. A. Silva, …, D. Baker. 2017. Massively parallel *de novo* protein design for targeted therapeutics. *Nature.* 550:74–79.

10. McCully, M. E., and V. Daggett. 2012. Folding and dynamics of engineered proteins. *In* Protein Engineering Handbook. S. Lutz and U. T. Bornscheuer, eds. Wiley-VCH, pp. 89–114.

11. Dahiyat, B. I., and S. L. Mayo. 1997. De novo protein design: fully automated sequence selection. *Science.* 278:82–87.

12. Marshall, S. A., and S. L. Mayo. 2001. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* 305:619–631.

13. Marshall, S. A., C. S. Morgan, and S. L. Mayo. 2002. Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* 316:189–199.

14. Shah, P. S., G. K. Hom, …, S. L. Mayo. 2007. Full-sequence computational design and solution structure of a thermostable protein variant. *J. Mol. Biol.* 372:1–6.

15. Gillespie, B., D. M. Vu, …, K. W. Plaxco. 2003. NMR and temperature-jump measurements of *de novo* designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. *J. Mol. Biol.* 330:813–819.

16. Mayor, U., C. M. Johnson, …, A. R. Fersht. 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA.* 97:13518–13522.

17. Mayor, U., N. R. Guydosh, …, A. R. Fersht. 2003. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature.* 421:863–867.

18. McCully, M. E., D. A. Beck, and V. Daggett. 2013. Promiscuous contacts and heightened dynamics increase thermostability in an engineered variant of the engrailed homeodomain. *Protein Eng. Des. Sel.* 26:35–45.

19. Eswar, N., B. Webb, …, A. Sali. 2006. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinforma.* Chapter 5:Unit 5.6.

20. Clarke, N. D., C. R. Kissinger, …, C. O. Pabo. 1994. Structural studies of the engrailed homeodomain. *Protein Sci.* 3:1779–1787.

21. Religa, T. L. 2008. Comparison of multiple crystal structures with NMR data for engrailed homeodomain. *J. Biomol. NMR.* 40:189–202.

22. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

23. MacKerell, A. D., Jr., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

24. Mackerell, A. D., Jr., M. Feig, and C. L. Brooks, III. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

25. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

26. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

27. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1998. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926.

28. Beglov, D., and B. Roux. 1998. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* 100:9050.

29. Beck, D. A. C., M. E. McCully, …, V. Daggett. 2000–2019. In Lucem Molecular Mechanics (ilmm). University of Washington, Seattle, WA.

30. Doreleijers, J. F., A. J. Nederveen, …, E. L. Ulrich. 2005. BioMagResBank databases DOCR and FRED containing converted and filtered sets of experimental NMR restraints and coordinates from over 500 protein PDB structures. *J. Biomol. NMR.* 32:1–12.

31. Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79:5333–5339.

32. Shire, S. J., G. I. Hanania, and F. R. Gurd. 1974. Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. *Biochemistry.* 13:2967–2974.

33. Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

34. Ibarra-Molero, B., V. V. Loladze, …, J. M. Sanchez-Ruiz. 1999. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry.* 38:8138–8149.

35. Strickler, S. S., A. V. Gribenko, …, G. I. Makhatadze. 2006. Protein stability and surface electrostatics: a charged relationship. *Biochemistry.* 45:2761–2766.

36. Contessoto, V. G., V. M. de Oliveira, …, V. B. P. Leite. 2018. TKSA-MC: a web server for rational mutation through the optimization of protein charge interactions. *Proteins.* 86:1184–1188.

37. Wang, J., R. J. Oliveira, …, V. B. Leite. 2012. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. USA.* 109:15763–15768.

38. Whitford, P. C., J. K. Noel, …, J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins.* 75:430–441.

39. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

40. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

41. Noel, J. K., M. Levi, …, P. C. Whitford. 2016. SMOG 2: a versatile software package for generating structure-based models. *PLoS Comput. Biol.* 12:e1004794.

42. Oliveira, R. J., P. C. Whitford, …, J. Wang. 2010. Coordinate and time-dependent diffusion dynamics in protein folding. *Methods.* 52:91–98.

43. Querol, E., J. A. Perez-Pons, and A. Mozo-Villarias. 1996. Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* 9:265–271.

44. Vogt, G., S. Woell, and P. Argos. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* 269:631–643.

45. Alford, R. F., A. Leaver-Fay, …, J. J. Gray. 2017. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* 13:3031–3048.

46. Bolon, D. N., and S. L. Mayo. 2001. Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry.* 40:10047–10053.

47. Bowman, G. R., and P. L. Geissler. 2014. Extensive conformational heterogeneity within protein cores. *J. Phys. Chem. B.* 118:6417–6423.

48. Dantas, G., B. Kuhlman, …, D. Baker. 2003. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* 332:449–460.

49. Hollien, J., and S. Marqusee. 1999. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry.* 38:3831–3836.

50. Schuler, B., W. Kremer, …, R. Jaenicke. 2002. Role of entropy in protein thermostability: folding kinetics of a hyperthermophilic cold shock protein at high temperatures using ^{19}F NMR. *Biochemistry.* 41:11670–11680.