PROTEINS: Structure, Function, and Bioinformatics 57:128–141 (2004)

An Electrostatic Basis for the Stability of Thermophilic Proteins

Brian N. Dominy,[1] Hervé Minoux,[2] Charles L. Brooks III[3]*

[1]*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts*
[2]*Aventis Pharma 13, Quai Jules Guesde, 94400 Vitry-Sur-Seine, France*
[3]*Department of Molecular Biology, TPC6, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California*

ABSTRACT **Two factors provide key contributions to the stability of thermophilic proteins relative to their mesophilic homologues: electrostatic interactions of charged residues in the folded state and the dielectric response of the folded protein. The dielectric response for proteins in a "thermophilic series" globally modulates the thermal stability of its members, with the calculated dielectric constant for the protein increasing from mesophiles to hyperthermophiles. This variability results from differences in the distribution of charged residues on the surface of the protein, in agreement with structural and genetic observations. Furthermore, the contribution of electrostatic interactions to the stability of the folded state is more favorable for thermophilic proteins than for their mesophilic homologues. This leads to the conclusion that electrostatic interactions play an important role in determining the stability of proteins at high temperatures. The interplay between electrostatic interactions and dielectric response also provides further rationalization for the enhanced stability of thermophilic proteins with respect to cold-denaturation. Taken together, the distribution of charged residues and their fluctuations have been shown to be factors in modulating protein stability over the entire range of biologically relevant temperatures. Proteins 2004; 57:128–141.** © **2004 Wiley-Liss, Inc.**

Key words: cold shock protein (Csp); chemotaxis Y (CheY); dielectric response; thermostability; thermophile; mesophile; hyperthermophile

INTRODUCTION

Life exists in virtually every environment on earth. Organisms have evolved to cope with many conditions, including extremes of temperature, salinity, acidity and pressure.[1,2] Often the proteins in these organisms are specially designed to function in the harsh environments in which they live. Research to identify the molecular origin of extremophilic adaptations is still in its infancy. However, significant biochemical and structural work has emerged about the specific class of temperature-resistant adaptations found in thermophilic and hyperthermophilic proteins.[3,4,5] Understanding the physical connection between sequence differences in mesophilic and thermophilic proteins and the tolerance of these molecules for high temperature is a first step toward understanding the robust nature of life.

Multicellular animals and plants generally do not tolerate temperatures over 323K (50°C). However, some organisms demonstrate optimal growth close to and even beyond this limit, for example, at temperatures above 318K (45°C) for thermophiles and above 353K (80°C) for hyperthermophiles.[2,6,7] Since the discovery of these extremophile families, different approaches have been undertaken to determine the physicochemical origins of the enhanced stability of proteins expressed by these organisms.[3,4,8–10]

Understanding the origin of protein thermostability is motivated by two goals. First, many thermophilic proteins are derived from ancient species and are believed to be related to proteins from more common mesophilic organisms.[4] Elucidating the physical properties differentiating these protein classes improves our understanding of evolutionary processes. Second, a successful design method for thermophilic enzymes would have enormous implications for industrial processes.[3] The reaction chemistry performed using protein catalysts is often limited by the relatively low optimal active temperature typical of most proteins. Some specific examples of thermophilic protein design are present in the literature;[11–15] however, a general and reliable approach has not yet been established. The ability to manipulate the sequence of a mesophilic protein in order to produce a thermophilic variant will have great promise in commercial applications.

Three-dimensional structures of several thermophilic and hyperthermophilic proteins have been elucidated.[4] Based on this structural data, B-factor analysis has suggested that thermophilic proteins may exhibit greater rigidity than mesophilic proteins.[16] Although the suggestion has been made that increased rigidity may generally differentiate mesophilic from thermophilic proteins, this has not yet been tested systematically.

Other structural features of thermophiles have also yielded hypotheses describing their enhanced stability. Shorter loop regions in some hyperthermophilic proteins has led to the suggestion that limiting fluctuations, and

*Correspondence to: Charles L Brooks III, Department of Molecular Biology, TPC6, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037. Email: brooks@scripps.edu.

Received 21 August 2003; Revised 28 November 2003; Accepted 12 December 2003

Published online 22 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20190

thus chain entropy in the folded and unfolded states, might generally lead to greater stability.[17] A greater density of hydrophobic groups in the protein core of some thermophiles has fueled the idea that packing may generally explain thermostability. The difficulty with each of these explanations is that they do not appear to generally describe stability in the thermophilic class of proteins. Despite the collection of data regarding individual crystal structures, no general explanation has been elucidated for the enhanced stability observed in thermophilic proteins.[3]

Analysis of nucleotide coding sequences from hyperthermophilic organisms has uncovered an increased number of charged amino acids and a reduced number of polar residues relative to proteins from mesophiles.[18,19] This trend has also been observed in a number of structural studies.[20,21] This motivates the suggestion that electrostatic interactions may play a crucial role in thermal stability.[22–24] However, the presence of enhanced surface charges has been discounted as a reason for the extraordinary stability of these systems because, based on simple applications of continuum electrostatic calculations, such interactions often only weakly stabilize, or destabilize, the folded protein.[25–27] As an alternative to electrostatic control, a rationalization for the thermal stability observed for extremophiles has been attributed to a combination of different factors, such as hydrogen bonding, hydrophobic packing, secondary structure propensity and helix dipole stabilization.[4]

Theoretical calculations and experimental analyses, however, suggest that small changes to protein surface charges, and thus electrostatic effects, can have dramatic consequences for protein stability. In recent studies by Schmid and coworkers,[28] Dominy and coworkers[22] and Dong and coworkers,[24] the differential dependence of mesophilic and thermophilic protein stability on salt strongly suggests the key role electrostatics play in enhancing the thermal stability of thermophilic and hyperthermophilic proteins. Another study by Elcock has shown that the addition of salt bridges, which generally destabilizes a protein at room temperature, may stabilize the same protein at higher temperatures.[29] This was attributed to the temperature dependence of the hydration free energy associated with charged residues. Furthermore, it has been shown experimentally by Grimsley and coworkers that single amino acid substitutions to charged surface residues can enhance protein stability.[30] Finally, in another study it is noted that the placement of charged residues may be optimized in hyperthermophilic proteins, with regard to their mesophilic homologues, to yield more favorable electrostatic interactions.[25] These results are in accord with the observed increase in the number of salt bridges found in thermophilic and hyperthermophilic proteins versus those found in their related mesophilic forms.[4,31] However, a simple increase in the number of ionic pairs does not explain the enhanced thermal stability of all hyperthermophilic proteins, as noted by Usher and coworkers.[32] This becomes particularly clear from hydrogen exchange experiments described by Hollien and Marqusee,[10] who show that enhanced thermal stability in a variant of RNase H from *Thermus thermophilus* arises from increased stability throughout the structure and is not localized to the specific regions of sequence differences.

In addition, work carried out by Lebbink and coworkers has demonstrated that mutations consistent with higher thermostability are not additive.[12] They investigated the effect of mutations altering glutamate dehydrogenase (GDH) from *Thermatoga maritima (Tm)* to cause it to more closely resemble the homologous enzyme from *Pyrococcus furiosus* (*Pf*), an even more thermostable protein. The mutation of three polar residues in *Tm* GDH to the corresponding charged residues in *Pf* GDH was found to be non-additive, and it was suggested that a long-range cooperative effect must be involved. This global, cooperative effect is consistent with the results described by Hollien and Marqusee for RNAse H.

In this paper, we present results illustrating how the electrostatic properties of proteins can modulate their thermal stability. The theoretical framework we develop moves beyond simple consideration of the interactions of individual residues to include the global response of the protein and solvent mediating these interactions. It provides a basis from which an explanation of differences in stability observed at room temperature emerges, but it also yields a rationalization for the differential thermal stability of mesophilic, thermophilic and hyperthermophilic homologs over the entire biologically relevant temperature range. We believe this model is applicable to a large fraction of thermophilic proteins.

In the current study, we examine proteins from two structural families that exhibit a range of thermal stability differences between homologs. Three proteins from the cold shock protein (Csp) family, expressed by the mesophilic bacterium *Bacillus subtilis* (*Bs*),[33] the thermophilic bacterium *Bacillus caldolyticus* (*Bc*)[34] and the hyperthermophilic bacterium *Thermotoga maritima* (*Tm*), are investigated (Fig. 1).[35] Additionally, two chemotaxis Y (CheY) proteins are studied, those expressed by the mesophile *Bacillus subtilis*[36] and by the hyperthermophile *Thermotoga maritima*[32] (Fig. 1). The crystal structures of the mesophilic cold shock protein[37,38] and the hyperthermophilic CheY protein[32] were available at the time of this study and were used in our work. Motivated by the high degree of sequence homology to the known crystal structures, the remaining structures were modeled by homology.[39] In addition to the availability of structural models, these two protein families were chosen because they are functionally unrelated, have different chain lengths and possess different topologies. Therefore, we have no *a priori* reason to suspect that they will utilize similar mechanisms to achieve thermostability.

Molecular dynamics (MD) simulations were performed on these protein systems. From the resulting trajectories, we examined the differences in global fluctuations (hypothesized to differentiate mesophilic from thermophilic proteins) as well as the coupling of the charge model and dynamics through the dielectric response.

From our analysis, we find that global dynamic fluctuations do not vary significantly between mesophilic and

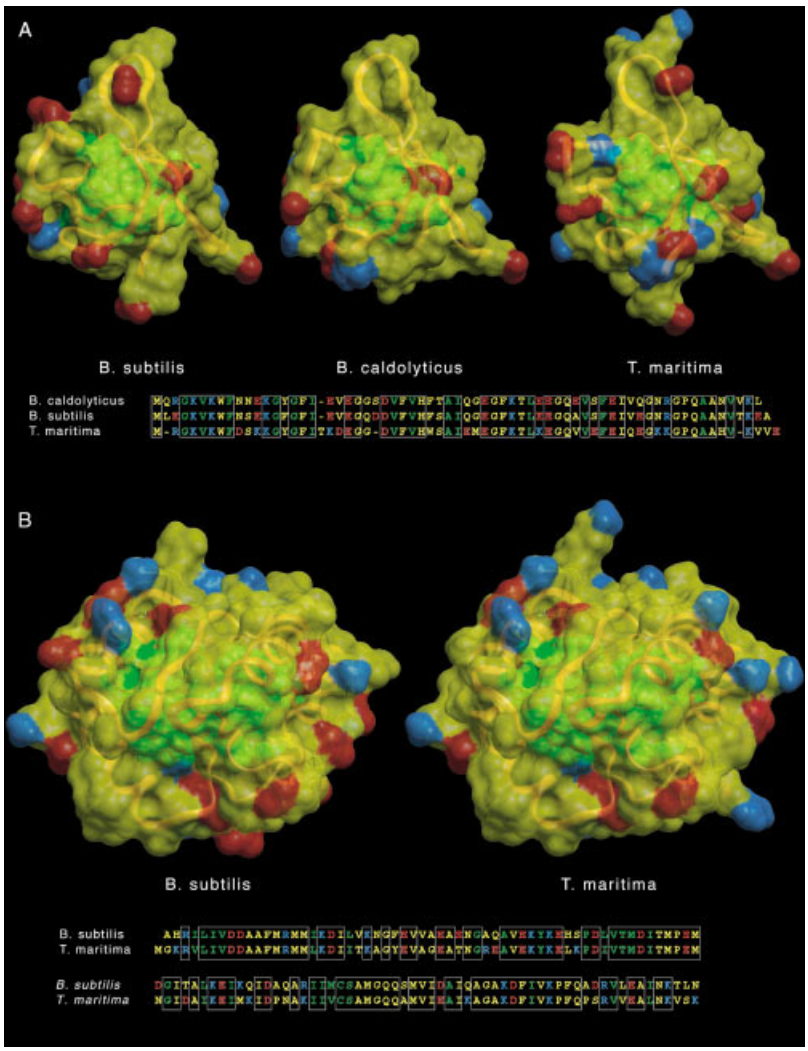

Fig. 1. Structure and sequence alignment for Csp (**A**) and CheY (**B**) proteins showing the conserved residues (boxed), the almost completely conserved hydrophobic core (green), and the differences in the distribution of positively (blue) and negatively (red) charged residues. Proteins within a family present a similar backbone (orange ribbon).

thermophilic proteins. In particular, no trend indicating increased rigidity with thermostability is observed. We do find, however, that the effect of dynamics on the charge distribution does vary with thermostability. Specifically, the dielectric response, which describes the fluctuation of the protein's dipole moment, increases with thermostability.

To explore the consequences of this observation for the stability of a protein, we have used continuum electrostatic models to compute the influence of a varying dielectric response on protein stability. The larger dielectric constant of thermophilic proteins may be linked to enhanced stability using these continuum models. Further, we find that the connection between the dielectric constant and protein stability is robust with regard to the implementation of the continuum model.

Finally, using this conceptual framework for how electrostatic properties determine thermostability, we designed two cold shock protein mutants. These mutants are found

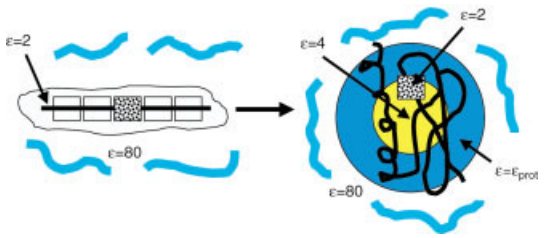

Fig. 2. A cartoon illustrating the multiple dielectric regions used when solving the Poisson equation to determine the electrostatic contribution of a given residue to protein stability. For the unfolded conformation (on the right), the neighboring regions of the residue in question are considered to be a low dielectric medium, while the external dielectric of the solvent is consistent with water. This is identical to most dielectric models of the unfolded state that utilize two dielectrics. The folded state (on the right) is modeled using the same low dielectric constant for the residue in question; however, the remainder of the protein is comprised of multiple dielectric regions. The core is considered to have a dielectric constant of 4, while the surface of the protein is assigned a higher dielectric constant. The solvent is still assigned a dielectric constant consistent with water at room temperature.

to exhibit dielectric response and electrostatic properties consistent with wild-type thermophilic species based on theoretical investigations described herein. This design protocol represents a first step toward a general and reliable approach for engineering protein thermostability.

MATERIALS AND METHODS

This section is organized as follows. First, we describe the systems used to explore the origins of enhanced thermostability in proteins. Next, we discuss the molecular dynamics protocols used to study the Csp and CheY protein families. We describe the calculation of the dielectric response for proteins as well as the Lindemann constant used to analyze protein flexibility. Finally, we elaborate the models used in our calculation of electrostatic contributions to protein stability.

Model Systems

The systems used in this study include three proteins from the Csp family and two CheY proteins, as noted above (Fig. 1). Homology models were built for the thermophilic and hyperthermophilic homologs in Csp and the mesophile in CheY. This strategy is reasonable because of the high degree of sequence identity to the known crystal structures. In the case of the Csp family the thermophile (*Bc*-Csp) and hyperthermophile (*Tm*-Csp) are 82% and 62% identical in sequence relative to the mesophile (*Bs*-CspB). Given this level of homology, structure modeling approaches are highly successful.[40] This claim was later validated when the crystal structure of *Bc*-Csp became available.[41] A comparison between the modeled structure and the newly available crystal structure showed small differences in the backbone conformation (<1 Å RMSD). In addition, an analysis of the electrostatic properties of the crystal structure demonstrated very similar results compared to the modeled structure and did not affect the conclusions of this study. The homology modeling package InsightII (Accelrys) was used to place the mutated side chains into the corresponding crystal structure. These side chains were refined by minimizing in the context of the rest of the protein, which was kept in the crystal conformation. The purpose of this minimization was simply to remove any van der Waals clashes and optimize local electrostatic interactions.

Molecular Dynamics Simulations

For each of the Csp proteins, two sets of MD simulations were carried out. One used an explicit representation of the solvent comprising 3796, 3800 or 3803 TIP3P water molecules surrounding the protein.[42] The simulation system consisted of the central cell with a truncated octahedron shape and full periodic boundary conditions. The force field used was the all hydrogen force field, param22, from CHARMM.[43] Coulombic energies were determined using a Particle-Mesh Ewald technique, and both temperature and pressure were kept constant at 25°C and 1 bar, respectively. A second set of MD simulations was carried out, also at 25°C, using the generalized Born (GB) model[44] to implicitly take into account solvent effects. These simulations were carried out using the CHARMM param19 force field.[45,46] The GB model, in the context of the param 19 force field, has been shown to reliably replace explicit solvent with regard to average atomic fluctuations while dramatically reducing computational cost.[44] These two sets of calculations were conducted using the CHARMM molecular mechanics package.[47] The SHAKE algorithm[48] was used in all simulations to constrain the H-heavy atom bond lengths and the time step was 2 fs. An equilibration time of 200 ps for *Bs*-CspB and of 600 ps for *Tm*-Csp and *Bc*-Csp was followed in all cases by simulations to 2 ns. A longer equilibration time for *Bc*-Csp and *Tm*-Csp was used in order to allow these proteins to relax (recall these systems were derived from homology built models). Both the explicit solvent and GB models were used to calculate dielectric constants of the Csp family of proteins. The similarity in the dielectric constants predicted using GB and explicit solvent simulations (differing an average of 12%) justified the use of the GB model on the CheY family (Table I). An implicit solvent representation greatly facilitated this study, since the mesophilic and hyperthermophilic CheY proteins (119 and 120 residues respectively) are approximately twice the size of the mesophilic, thermophilic and hyperthermophilic proteins from the Csp family (66, 67, and 66 residues respectively).

TABLE I. Values of Dielectric Constant from Molecular Dynamics and Normal Mode Analysis

	Protein Dielectric Constants		
Model	TIP3P	GB	Normal Mode
Bs-CspB	12	13	5
Bc-Csp	16	18	4
Tm-Csp	25	30	11
Bs-CheY	N/A	11	3
Tm-CheY	N/A	20	4

The dielectric constants computed from a normal mode analysis are smaller than those obtained from the MD simulations because only a portion of the total flexibility is accounted for in the normal mode calculations. Structures were minimized and analyzed under identical force field conditions.

Dielectric Response Via Molecular Dynamics

The dipole fluctuations of the atomic and simplified protein models were analyzed in the context of the Fröhlich–Kirkwood theory to determine the intrinsic dielectric constant for the protein systems we studied.[49] The approach is the same as that employed by Simonson and Brooks in studying a variety of mesophilic proteins.[49] The expression for the dielectric constant from the Fröhlich–Kirkwood theory is related to the so-called *G*-factor, which is proportional to the average square fluctuation of the molecular dipole moment and inversely proportional to the temperature:

$$G = \frac{(2 \cdot \varepsilon_w + 1)(\varepsilon_p - 1)}{(2\varepsilon_w + \varepsilon_p)} = \frac{\langle \Delta M_P^2 \rangle}{k_B T r_p^3} \quad (1)$$

where q_i is the partial charge of atom *i,* m_i is the mass of atom *i,* ε_w is the dielectric constant of the solvent, ε_p is the

dielectric constant of the protein and r_p is the approximate spherical radius of the protein (16 Å for the Csp and CheY families). It is worthwhile at this juncture to comment on the anticipated behavior of the *G*-factor with protein fluctuations. In the harmonic limit, the mean-square fluctuation of an atom is proportional to temperature and therefore the contribution to the dielectric constant is invariant to changes in temperature. This can be seen by considering the expression for ΔM_p^2, as noted in the equation below [Eq. (2)].

$$G = \frac{\langle \Delta M_P^2 \rangle}{k_B T r_p^3} = \frac{\left\langle \left[\sum_i^{Ncharges} q_i(\mathbf{r}_i - \langle \mathbf{r}_i \rangle) \right]^2 \right\rangle}{k_B T r_p^3} \sim \frac{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle}{k_B T r_p^3} \overset{\text{Harmonic Limit}}{\propto} \frac{(\mathrm{k_B T/m_i})}{\mathrm{k_B T r_p^3}} \quad (2)$$

We note that this global continuum description of the dielectric constant for a protein provides a connection between the microscopic motions of charges in the protein and the macroscopic response of the protein to changing electric fields. However, it is not the only means of describing a protein's response to either internal or external changes in electric fields. More microscopic models, as have been described by Warshel and coworkers,[50,51] begin to account for the localized response of the protein to changes in the charge distribution. These models, presumably, capture more accurately the detailed energetic aspects of 'dielectric response', but with a significant cost in complexity and physical insight.

Atomic Fluctuations and Lindemann Constant

The Lindemann parameter provides a global description of atomic fluctuations in a given protein system.[52] Values of the Lindemann parameter were evaluated from the MD trajectories, using an explicit (TIP3P) or implicit (GB) solvent, for the Csp and CheY proteins. The Lindemann parameter for protein atoms within a sphere of radius r_{cut}, centered at the centroid of the protein is obtained using the following expression:[52]

$$\Delta_L^{in}(r_{cut}) = \frac{\sqrt{\sum_{i,r<r_{cut}} \langle \Delta r_i^2 \rangle / N}}{a'} \quad (3)$$

Where N is the number of atoms within the spherical region, a' the most probable non- bonded near-neighbor distance (typically 4.5 Å and so defined in this study),[52] and $\Delta r_i^2 = (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2$, with $\mathbf{r}_i$ the position of atom *i*.

Calculating Electrostatic Stability

The GB Model

The electrostatic contribution to the free energy of folding was calculated from the *Bs*- CspB and *Tm*-CheY crystal structures and previously described homology models. These structures were minimized under restraints within the CHARMM/GB param19 force field[44] and used as models of the folded state. The GB term within the force field was parameterized using an approach based on that of Qiu et al.[53] The unfolded state for each protein was modeled as a *trans* peptide chain with side chains also built in a *trans* conformation in which the i^{th} residue only interacts with $i \pm 2$ residues (its nearest neighbors). Other models have also been developed for computing electrostatic contributions to the folding free energy.[25,26] In these cases, only side chain interactions are considered, while backbone interactions are represented implicitly. For the purposes of this study, we were interested in constructing a model that provides a complete description of the electrostatic free energy difference upon protein folding, keeping in mind that sequential short range electrostatic interactions may still be present in the unfolded state. In this model, the total electrostatic energy has converged with respect to the window size of adjacent residues and is therefore robust in describing the energetics of a broad range of structural models of the unfolded state, assuming that such states are unstructured random coils. The same model has been used successfully in a continuum electrostatic study of ionic strength effects in the Csp family.[22] In this study, electrostatic contributions of individual residues to protein stability were found to agree well with experiments carried out on the same protein family.

The electrostatic folding free energy was computed as a function of the protein dielectric constant using the following expressions:

$$\Delta G_{\text{folding}}^{\text{elec}} = \frac{1}{\varepsilon_p}\left(E_{\text{folded}}^{\text{Coulomb}} - E_{\text{unfolded}}^{\text{Coulomb}} \right) + \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right)\left(E_{\text{folded}}^{\text{GB}} - E_{\text{unfolded}}^{\text{GB}} \right)$$

with

$$E_{\text{(un)folded}}^{\text{GB}} = \frac{1}{2} \sum_i \sum_j \frac{-q_i q_j}{\{r_{ij}^2 + \alpha_i \alpha_j \exp[-r_{ij}^2/(4\alpha_i \alpha_j)]\}^{1/2}} \quad (4)$$

where the Born radii[44] (α_i) and atomic pair distances (r_{ij}) were computed within the corresponding folded and unfolded structural models described previously. The Born radius may be thought to represent an average distance between the atom and the solvent boundary, or the degree of burial for the atom. The Born radii calculated for the unfolded model are closely related to the van der Waals radii, reflecting greater solvent exposure. This continuum-based model provides a simple connection between the sequence differences associated with proteins in a 'thermophilic series', the observed differences in the dielectric response of each protein and the differences in stability of these proteins. Semi-microscopic models that attempt to partition the protein response between localized structural relaxation and the overall protein response replace this simple interpretation of the protein dielectric response with a more localized description of energetic changes,[51] such as may be important in reproducing pK_a values compared with experiment. However, these models can obscure more global aspects of a phenomenon such as

the one we are exploring here and are unnecessary to provide the conceptual basis for our understanding.

Multiple Dielectric Poisson–Boltzmann Models

To explore more detailed models of electrostatic folding stability, multiple dielectric grid- based Poisson–Boltzmann (PB) models were also utilized in calculating electrostatic stability via numerical finite difference methods.[54] Using this approach, individual atoms or groups of atoms could be individually assigned a dielectric constant. This 'painting' of the dielectric response to various regions of the protein permitted us the greatest flexibility in designing multiple dielectric models of the folded protein. These more complex continuum electrostatic models allowed us to explore the model dependence of the electrostatic stability calculations performed with the two dielectric GB continuum approach.

The models of the folded and unfolded states of the proteins under study were similar to those used in the two dielectric GB approach. The only difference here was that, instead of modeling the unfolded state as a linear chain, each residue interacts with $i \pm 2$ sequential residues in the folded conformation. In this way, local, sequential interactions completely cancel from the unfolded to the folded state. The major differences in the model compared to the model encompassed by Eq. (4) are as follows:

- The use of the Poisson equation to determine the electrostatic free energy rather than a sum of the generalized Born and Coulomb equations.
- The use of multiple solute dielectric regions as opposed to a single dielectric constant to represent the folded and unfolded protein interior.

Furthermore, within the context of our PB-solver, a van der Waals surface was used to define multiple dielectric regions. The use of a van der Waals surface as opposed to the solvent inaccessible or molecular surface[55–57] enhances the electrostatic free energy of solvation because the solvent boundary is effectively closer to the solute. This is primarily due to the lack of re-entrant surfaces included in the solvent inaccessible surface. Nevertheless, the effect on the quantitative results is small, and no change is observed regarding the qualitative trends when compared to the generalized Born model. This is demonstrated by the good agreement found between the two dielectric finite difference Poisson-Boltzmann (FDPB) results and the results obtained using a two dielectric generalized Born model (see discussion below). Therefore, we believe the multiple dielectric models provide an appropriate complement to the simple generalized Born theory just described [Eq. (4)].

The first implementation of this FDPB approach is analogous to the two dielectric model used in the GB calculations described above. In the two dielectric model, the dielectric regions are assigned similarly to the way they are in the GB approach, where the protein interior in the folded and unfolded state is assigned a single dielectric constant. The only difference here is the slightly different dielectric boundary defined by the molecular surface. As described earlier, the stability calculation as a function of the internal dielectric is shown to demonstrate very similar trends to the GB model for both the Csp and CheY systems. Following the same mechanics demonstrated with the two dielectric GB model, the stability calculations are performed as a function of the protein dielectric constant (varying between a dielectric constant of 1 and 80). The discussion will highlight the changes in electrostatic stability due to changes in the dielectric constants as determined by the Fröhlich–Kirkwood equation for mesophilic and thermophilic species within the Csp and CheY protein families.

In the three dielectric model, the internal dielectric constants of an individual residue and the surrounding $i \pm 2$ residues are assigned a value of 2, consistent with small molecule studies of solvation free energies.[58] In order to calculate the solvation energy of this residue (i) in the unfolded state ($\Delta G^{\text{solv}}_{\text{i,unfolded}}$), the reaction field energy is computed for the individual residue surrounded by its (uncharged) nearest neighbors and absent the remainder of the protein, where the dielectric constant for the surrounding bulk solvent is assigned a value of 80.

In addition, the electrostatic potential generated by the charges on neighboring $i \pm 2$ residues at the positions of the charges in the residue i may be computed. Multiplying the potential by the charges on the residue results in the electrostatic interaction energy between the residue and its sequential neighbors in the unfolded state ($\Delta G^{\text{interaction}}_{i,\text{ unfolded}}$).

In the folded state model, the residue in question is still surrounded by the local low dielectric of 2 assigned to itself and its $i \pm 2$ nearest neighbors. The remainder of the protein has a higher dielectric constant. In the case of the Csp mesophile, this dielectric constant would correspond to 12, as shown in Table I. As in the two dielectric GB model, the electrostatic stability of the protein is calculated as a function of the bulk protein dielectric constant. The protein itself is surrounded by a dielectric of 80, representing water that extends indefinitely, thereby assuming infinite dilution of the protein solute. A reaction field, or solvation, energy may then be computed for the residue in question in the folded protein environment ($\Delta G^{\text{solv}}_{i,\text{ unfolded}}$). The difference in this solvation energy and that in the unfolded state is the electrostatic penalty involved in burying the residue.

The electrostatic potential field generated by the protein at positions consistent with the individual residue charges may also be computed in this dielectric environment. The residue charges, multiplied by the electrostatic potential from the protein, result in the interaction energy between the protein and the residue:

$$\Delta G^{\text{interaction}}_{i,\text{ unfolded}} = \sum_{k}^{\text{Ncharges}} q^i_k \phi_k \tag{5}$$

where q^i_k refers to a charge q at site k located on residue i and ϕ_k refers to the electrostatic potential at site k due to the surrounding charges.

The same approach may be used to compute the interaction between the residue in question and the nearest-neighbor residues. The sum of the interaction terms yields the total electrostatic interaction between the residue and its environment within the folded protein ($\Delta G_{i,\,\mathrm{unfolded}}^{\mathrm{interaction}}$). The difference between the solvation and interaction terms in the folded and unfolded states results in the contribution of a given residue to the electrostatic free energy of protein stability.

$$\Delta G_{\mathrm{folding}}^{\mathrm{elec}} = \sum_{i}^{\mathrm{Nres}} (\Delta G_{i,\mathrm{folded}}^{\mathrm{solv}} - \Delta G_{i,\mathrm{unfolded}}^{\mathrm{solv}}) + \frac{1}{2}(\Delta G_{i,\mathrm{folded}}^{\mathrm{interaction}} - \Delta G_{i,\mathrm{unfolded}}^{\mathrm{interaction}}) \quad (6)$$

The four dielectric model for electrostatic stability simply builds on the three dielectric approach. Here, the folded protein is separated into two regions based on their distance from the center of geometry. In the case of Csp and CheY, an arbitrary distance of 8 Å was chosen to represent the boundary between the protein core and the protein surface. The protein core was kept at a dielectric constant of 4, while only the exterior surface dielectric was variable. The electrostatic stability was then determined as a function of the surface dielectric response. A schematic of the protocol used to compute the per-residue contributions to protein stability is shown in Figure 2.

RESULTS AND DISCUSSION

In this section, we present and discuss results from our calculations for both the Csp and CheY families of proteins. First we describe the correlation between dielectric response and thermostability. The dielectric response is shown to be closely connected to the number of charges present on the protein surface, a property also correlated with thermostability. In the Appendix, we illustrate the generality of this connection by using a simplified protein model and a large variety of charge distributions. We then demonstrate that the increased dielectric response, as well as the charge complementarity observed in thermophiles leads to greater stability. This point is demonstrated using a simple GB continuum electrostatic model. However, we also show these results to be independent of the methodological details through an examination of multiple dielectric Poisson calculations. We then address the issue of cold-denaturation and demonstrate that our model also rationalizes the enhanced stability of thermophiles in the presence of extreme cold. This furthers the claim that cold-denaturation could have a significant electrostatic basis, in addition to the well-known contributions from the hydrophobic component. Finally, we show results from the design of novel Csp mutants created to exhibit electrostatic properties similar to those of thermophilic and hyperthermophilic wild-type species.

Dielectric Behavior and Fluctuations of Extremophilic Proteins

Quantifying the dielectric response of the proteins alone yields an interesting trend. The dielectric constant is calculated through the Fröhlich–Kirkwood equation [Eq. (1)].[49] Using this equation, one can directly relate the fluctuations of the molecular dipole moment computed from MD simulations to the dielectric response. The calculated value of the dielectric constant for proteins from the Csp and CheY families (see Table I) is found to increase with the experimentally observed thermal stability. The convergence of the dielectric response during the course of the 2 ns simulations is shown in Figure 3. From our comparisons of both explicit-solvent, all-hydrogen force fields (param22 with explicit solvent) to those including only polar hydrogen atoms (param 19) with an implicit solvent, the dielectric constant is found to be independent of both the detailed nature of the protein force field and the solvent representation, suggesting it is a robust feature of the proteins themselves.

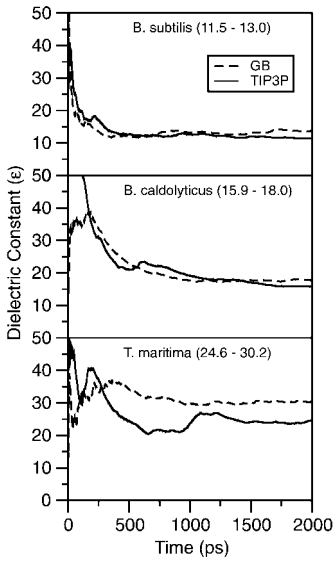

Fig. 3. Evolution of the dielectric constant during the course of the molecular dynamics simulations for proteins from the Csp family. The implicit solvent (GB) and explicit solvent (TIP3P) simulations produce similar results for the dielectric response of the three proteins shown above.

The probability distribution of the instantaneous Kirkwood *G*-factor is illustrated in Figure 4 and is found to agree well in all cases with the distribution predicted from continuum theory. Given the average *G*-factor value from the MD simulations, the continuum theory predicts dielectric constants very similar to those based on averaging over simulation snapshots. This demonstrates a level of convergence of the calculated dielectric constants consistent with the qualitative observation that the dielectric response appears to be correlated with thermostability.

For both the Csp and CheY families, thermal stability is correlated with an increase of the dielectric constant, which is a global property of the molecule. This observation is consistent with equilibrium hydrogen exchange studies of RNase H, carried out by Hollien and Marqusee.[10] In these studies, it was demonstrated that the general distribution of stability in the mesophilic and thermophilic forms of RNase H is similar, with only a

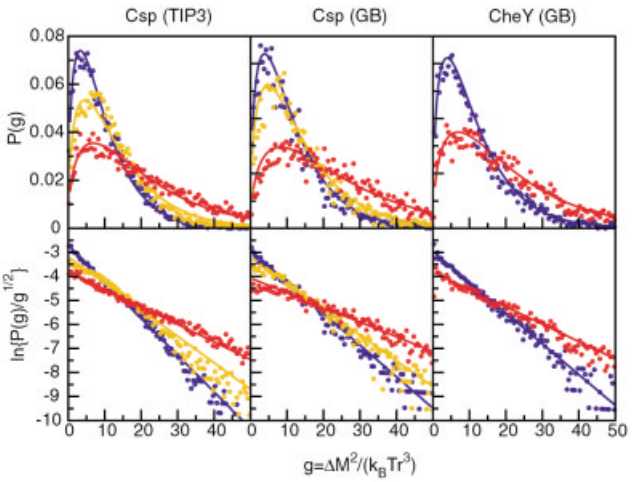

Fig. 4. Probability distribution functions of the Kirkwood g-factor from simulation and continuum theory. Theoretical curves are shown in solid lines while circles represent the *g*-factor values sampled from molecular dynamics simulations. Blue corresponds to the mesophilic proteins, orange to the thermophilic proteins, and red to the hyperthermophilic proteins.

proportional increase for all residues in the thermophile.[10] This proportional stability increase is correlated with our finding that the global protein dielectric response also increase with thermostability.

Our finding is also consistent with studies performed by Lebbink et al., where it was determined that long-range cooperative effects augment electrostatic interactions in a hyperthermophilic glutamate dehydrogenase (GDH).[12] It was also suggested by Lebbink et al. that the non-additive nature of the stabilizing interactions could result from the electrostatic fields generated by the involved residues in GDH. The dielectric response is a function of the number and distribution of charges contained within a protein and directly impacts long-range electrostatic interactions. The accumulation of salt-bridge networks within a protein can affect the system's stability, cooperatively, or non- additively through an increased Coulombic energy and a concomitant reduction of the desolvation penalty through the observed increase in the dielectric response. The connection between the increased dielectric response observed in thermophilic proteins and enhanced stability will be discussed in more detail later.

A thorough discussion of cooperativity in electrostatic interactions within thermophilic proteins is given in Karshikoff and Ladenstein.[5] This article explores the source of thermophilic protein stability and specifically the role of ion pairs in thermostability. An explanation of the non-additive or cooperative effects of ion pairing in thermophilic proteins observed in experiments[12,23] is described as involving conformational changes that produce a "more favourable polar environment." This description is consistent with an enhanced dielectric response, reflecting the ability of the charges within the protein to rearrange and stabilize the incorporation of a new charge. The systems described within this manuscript illustrate the cooperative mechanism described by Karshikoff and Ladenstein.

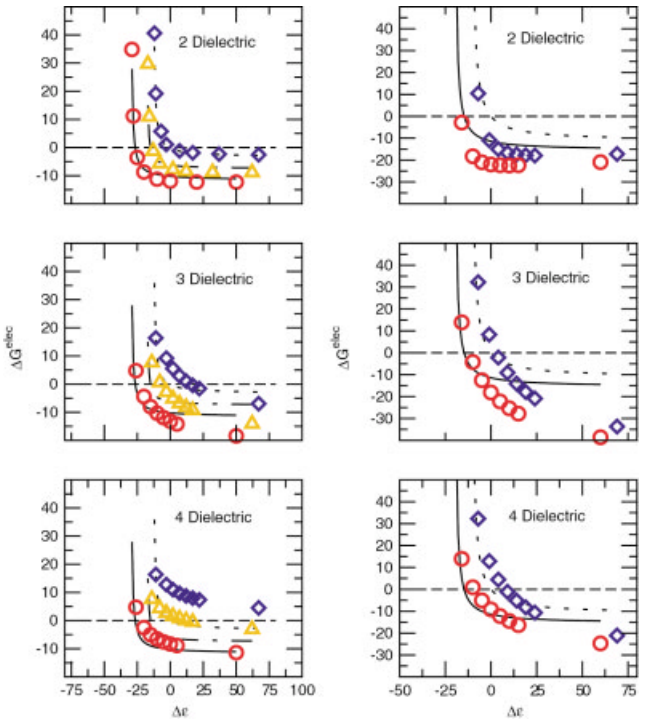

Fig. 6. Electrostatic component of the folding free energy (ΔG^{elec} in kcal/mol) is plotted as a function of the change in the interior dielectric response, $\Delta\varepsilon_p = \varepsilon_p - \varepsilon_p(298K)$. These curves correspond to the generalized Born model calculations shown in Figure 4. The symbols correspond to the results computed using the CHARMM finite difference Poisson–Boltzmann solver. Multiple dielectric models were used in these calculations. In the two dielectric models, the stability is calculated as a function of a changing dielectric constant representing the entire protein. In the three dielectric models, individual residues are restricted to a dielectric constant of 2, consistent with the proposed dielectric constant for a small molecule. In the four dielectric model, the protein core dielectric constant is fixed at 4, while only the dielectric at the protein surface is variable. Diamonds indicate the mesophilic proteins (*BS*-CspB, *Bs*-CheY), triangles indicate the thermophilic protein (*Bc*-Csp) and the circles indicate the hyperthermophilic proteins (*Tm*-Csp, *Tm*-CheY).

The dielectric response of a protein is related to its flexibility and charge distribution through the molecular dipole moment.[49] It is therefore important to distinguish which of these factors plays the dominant role in the observed differences between the dielectric constants within a 'thermophilic series.' To determine whether changes in protein flexibility at room temperature are correlated with the calculated changes in the dielectric constant, Lindemann parameters[52] for the Csp and CheY proteins have been determined using the relationship noted previously [Eq. (3)]. The Lindemann parameter provides a global indication of flexibility of the protein. From Table II we see that there are no significant differences in this property between the mesophilic and thermophilic proteins, suggesting no correlation between thermostability and flexibility. Slightly lower flexibility is indicated by the Lindemann parameter for the hyperthermophilic protein *Tm*-Csp using an explicit solvent model. However for both solvent models and for both the Csp and CheY systems, these small differences in flexibility cannot account for differences in the dielectric constant.

TABLE II. Values of Lindemann Parameter Evaluated From MD Trajectories, Using Explicit (TIP3P) or Implicit (GB) Solvent, for Csp and CheY Proteins

	Lindemann Parameters	
Model	TIP3P	GB
Bs-CspB	0.29	0.30
Bc-Csp	0.29	0.29
Tm-Csp	0.27	0.35
Bs-CheY	N/A	0.26
Tm-CheY	N/A	0.27

The higher dielectric constant for hyperthermophilic Csp and CheY appears to arise from the addition of complementary pairs of charged residues to the solvent accessible surface. The altered charge distribution increases the fluctuations of the molecular dipole moment, resulting in a higher dielectric constant despite similar thermal fluctuations. This behavior can be understood by considering the Fröhlich–Kirkwood equation for the dielectric constant [Eq. (1)]. As the charge density increases for a fixed magnitude of atomic fluctuations, larger mean-square fluctuations of the molecular dipole moment become possible [Eq. (2)]. Thus, the dielectric response of a protein is modulated by the distribution of charges on its surface.

Electrostatic Contributions to Thermophilic Protein Stability

Enhanced electrostatic stability is not derived solely from an increased dielectric response, although it is sufficient on its own to provide stability differences. The complementarity of the charge distribution also has a dramatic influence on stability by increasing favorable Coulombic interactions in the folded state. In fact, more favorable Coulombic interactions in the folded state may also be sufficient to account for differences in thermal stability, if the dielectric response of the protein is sufficiently large ($\varepsilon_p \geq 10$).[25] However, it is unclear that this factor alone can account for the global nature of stability differences that seem present within such systems, as demonstrated by Hollien and Marqusee,[10] or the cooperative stabilizing effect observed by Lebbink et al.[12] The total number of charged residues decreases in *Bc*-Csp relative to *Bs*-CspB, but the charge complementarity in *Bc*-Csp increases (Table III) and, correspondingly, the Coulombic contribution to stability is more favorable. The dielectric constant of *Bc*-Csp relative to *Bs*-CspB is also higher (this is true for *Tm*-Csp as well) due to the enhanced complementarity in the charge distribution. In the case of *Bc*-Csp, the changes in the charge distribution relative to *Bs*-CspB and their effects on the dielectric constant are not simply related to the number of charges, as in the hyperthermophilic proteins, but reflect subtler changes in the charge distribution, as might be expected given the closer relationship of *Bc*-Csp and *Bs*-CspB in terms of thermostability.

Developing networks of charged residues on the solvent accessible surface enhances thermal stability within the families of proteins from Csp and CheY. We have established that this can increase the intrinsic dielectric response of the protein, and also enhance the electrostatic stability of proteins within a 'thermophilic series.' To qualitatively illustrate this point, we have computed the electrostatic contribution to the folding free energy (ΔG^{elec}) as a function of the interior dielectric constant of the protein using a simple continuum based electrostatic model. We display these results in the form of 'corresponding states' curves in which the electrostatic contribution to folding stability is plotted versus the change in the dielectric constant of the protein, with each protein referenced to the value of its calculated dielectric constant at 298K (Fig. 5). When the protein dielectric constant is lower than its value at room temperature [$\Delta\varepsilon_p = \varepsilon_p - \varepsilon_p\,(298\text{K}) < 0$], the desolvation penalty outweighs the Coulombic interactions that favor the folded state; consequently, the net contribution to stability from electrostatic interactions is unfavorable. One factor that may distinguish mesophiles from thermophiles is the response they show to mutations that create or delete a salt-bridged neighbor. For proteins with small interior dielectric constants, the creation of salt-bridge interactions may be destabilizing, whereas analogous mutations in a thermophile will be stabilizing. Anecdotal experimental studies support this idea;[59–61] however, more systematic studies will have to be performed before general conclusions are reached.

This situation is consistent with past calculations, which have indicated that electrostatics will generally oppose protein folding when the protein dielectric constant is assumed to be small ($\varepsilon_p = 1$–4).[26,27] This is observed in all proteins from the Csp and CheY families. However, as this dielectric constant is increased, the electrostatic interactions begin to favor the folded state. This arises from a decreasing desolvation penalty, as the difference in the dielectric constant of the protein and solvent is diminished. Although the favorable Coulombic interactions are also diminished as the solute dielectric rises, these direct interactions are not attenuated as rapidly.

To summarize these key points, protein stability increases if the dielectric response of the protein increases (or that of the solvent decreases) and/or if the extent of favorable electrostatic interactions in the folded state is increased, which will generally lead to an increased dielectric constant for the protein. Thermophilic and hyperthermophilic proteins may be more stable, thermodynamically, than their mesophilic counterparts at a given temperature because of these two factors.

This is not the first time charge mobility has been suggested to contribute hyperthermophilic protein stability. In work by Aguilar et al., a crystal structure of a hyperthermophilic β-Glycosidase contains a large proportion of surface ionic pairs relative to the mesophilic homologue.[62] These authors argue that it is the high mobility and resilience of the salt bridge pairs that determines thermostability. In a separate work by Lebbink et al., the crystal structure of glutamate dehydrogenase from *Thermotoga maritima* was determined and found through *B*-factors to have highly mobile ionic pairs on its surface.[12]

TABLE III. Number of Charged Residues in Csp, CheY and Designed Mutant Proteins

	Distribution of Charged Residues						
	Bs-CspB	*Bc*-Csp	*Tm*-Csp	*Mutant A*	*Mutant B*	*Bs*-CheY	*Tm*-CheY
Arg	1	2	1	1	1	4	4
Lys	5	5	10	11	13	9	14
Asp	2	1	3	1	0	10	8
Glu	10	8	8	11	14	8	9
Total charge	−6	−2	0	0	0	−5	1
Total number of charges	18	16	22	24	28	31	35

The sequences for the designed mutants A and B are MLEGKVKWFNSEKGFGFIEVKGEKDVFVHFSAIEGEGFKTLKEGEAVSFEKVEGNRGPEAAKVTKKA and ELEGKVKWFNSEKGFGFIEVKGEKEVFVHFSAIEGEGFKELKEGEAVSFEKKEGNRGPEAAKVKKKA, respectively.

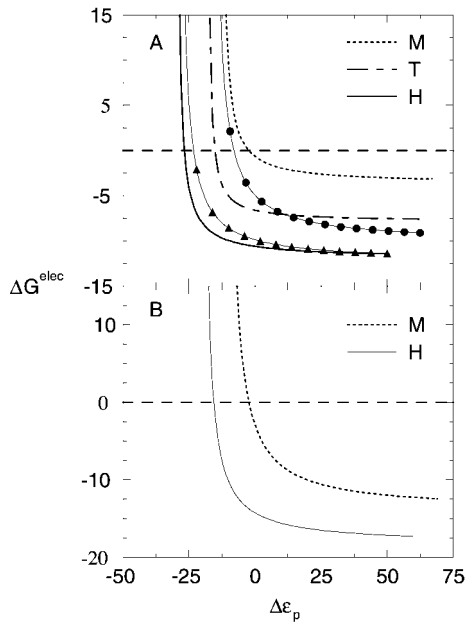

Fig. 5. Electrostatic component of the folding free energy (ΔG^{elec} in kcal/mol) is plotted as a function of the change in the interior dielectric response, $\Delta\varepsilon_p = \varepsilon_p - \varepsilon_p$(298K). Corresponding states stability curves for the proteins families (**A**) Csp and (**B**) CheY are *Bacillus caldolyticus* (dotted-dashed curve) and hyperthermophiles from *Thermotoga maritima* (solid curve). The stability curves are referenced to the value of dielectric constant occurring at 298K, thereby illustrating corresponding states within a family of homologues. Thermophiles and hyperthermophiles are generally more stable to changes in their interior dielectric constant, as might occur with changes in temperature.

Again, the suggestion was that the mobility of these charges was of key importance in determining the stability of this protein.

A question naturally arises as to whether the connection between enhanced stability and dielectric response is dependent on the details (or lack thereof) of the GB model. Further, while the dielectric response may be described macroscopically as a global property of the system, simulation results have shown that the surface charge residues play a leading role in determining the dielectric response in proteins.[49] This leads to a possible model in which the enhanced dielectric response observed in thermophilic proteins is limited to the surface of these proteins. In order to address both of these important concerns, an analysis was performed using multiple dielectric Poisson models. The results shown in Figure 6 demonstrate that all models show similar behavior. Even in the most restrictive case where the changing dielectric response is limited to the protein surface, a larger dielectric constant still leads to enhanced electrostatic stability.

In a recent paper by Mozo-Villarias et al., a "quasi-dipole moment" is shown to distinguish between mesophilic and thermophilic proteins in a significant fraction of the protein structures examined.[63] This is another global electrostatic property of proteins that is related to the dielectric constant studied in this work. In the cases studied in this paper (the Csp and CheY protein families) the dipole moment is observed to increase from mesophilic through hyperthermophilic proteins contrary to the general trend observed in the paper by Mozo-Villarias et al. The larger magnitude of the dipole moments observed in thermophilic proteins is further reflected in the larger fluctuations of the dipole moments, giving rise to larger dielectric responses. As such, the current work may reflect an alternative or complementary description for the electrostatic basis of thermostability. Further, the use of the dielectric constant suggests a sensitive connection between temperature and protein stability, a feature not fully explored in previous works.

Temperature Dependence of Protein Stability

We have demonstrated that adding charges to the molecular surface can enhance the dielectric response of a protein by allowing greater fluctuations of the molecular dipole moment. Temperature can also have an impact on the protein's dielectric response by motivating changes in atomic fluctuations and thereby fluctuations in the molecular dipole moment. The dielectric response of a protein changes with temperature (T) because of the relationship between the atomic fluctuations, which increase as T^{α} with $\alpha \geq 1$,[64] and the protein charge distribution. As the temperature is increased, the dielectric response increases due to contributions of anharmonic motion to atomic fluctuations of the protein. Harmonic motions do not contribute to the temperature dependence of the dielectric constant. Thus, we conclude that electrostatic interactions stabilize proteins at high temperatures, where their dielectric constant has increased due to increased atomic fluctuations. This arises from the diminution of the dielectric difference at the protein/water interface, occurring in part

from the reduction of the dielectric constant of water as temperature increases.[65]

Our model also suggests that these same electrostatic features may provide a rationalization for the reduced susceptibility of specific thermophiles and hyperthermophiles to cold-denaturation, as noted in previous experimental studies.[29,66,67] As the temperature is reduced, protein motion is eventually described by harmonic fluctuations (at approximately 220K).[68,69] The dielectric response of the protein now becomes independent of temperature and can be analyzed using a normal mode description.[70] In this harmonic limit, the value of the dielectric constant is significantly reduced from that observed in MD simulations at 298K and is closer to the value generally used in continuum solvation calculations (see Table I).[64] Therefore, we conclude that the dielectric constant, through the anharmonic contribution to atomic fluctuations, will decrease proportionally with decreasing temperature until approximately 220K, where only the temperature-independent, harmonic contribution will remain. Thus, until the temperature falls below approximately 220K, we can relate ΔG^{elec} versus $\Delta\varepsilon_p$ to ΔG^{elec} versus temperature, T. From this relationship and the results displayed in Figure 5, our model provides a consistent rationalization for differences in the cold-denaturation temperatures observed within 'thermophilic series.'

As the temperature decreases, the dielectric constant of the protein decreases toward a harmonic limit, and the protein is increasingly destabilized electrostatically (due to the electrostatic desolvation term that favors the unfolded state). At sufficiently low values of the dielectric constant, the desolvation term will dominate the favorable Coulomb interactions, cold-destabilizing the protein.[25–27] For mesophilic proteins relative to their thermophile homologues, the large derivative of the electrostatic folding free energy with respect to decreases in the protein dielectric constant near room temperature (e.g., for $\Delta\varepsilon_p < 0$) indicates that these proteins have a greater susceptibility to cold denaturation. We note that a temperature-independent dielectric constant qualitatively accounts for these differences in stability at low temperatures, but ignores the anticipated changes in atomic motions and their attendant manifestations on the protein's dielectric response just described. The importance of the contribution of electrostatics to cold-denaturation has also been previously noted.[71,72]

In Silico Designed Thermophilic and Hyperthermophilic Csp Proteins

Based on our model for electrostatic protein stabilization, we turned to the contribution of individual residues to protein stability. We were interested in whether our theoretical model would be able to reproduce a trend observed experimentally. It had been previously determined that two residues within the Csp thermophile, when mutated, have a significant impact on stability.[28] In fact, mutations at these positions can reduce the stability of the thermophile to that of the mesophile. In order to test our model, we created mutants *in silico* consistent with those used in the experimental work. For each *Bc*-Csp variant, the electrostatic contribution to the free energy of folding was calculated and compared to the one of the wild type, using a dielectric constant of 18 for both systems. The model identified R3E and L66E as the mutations most disruptive to the stability of *Bc*-Csp, exactly as was demonstrated experimentally.

Our observations and simple model analysis, as well as our success in identifying key residues associated with thermostability, led us to a design criterion for enhanced thermostability, as well as stability against cold-denaturation: optimization of the pairing and placement of complementary charged residues on the surface of a protein. To test this hypothesis, we designed variants of the mesophilic protein *Bs*-CspB by identifying surface-accessible residue pairs and replacing them by complementary pairs of charged residues (see Table III). In all, 10 residues were changed for mutant A and 15 residues for mutant B. The sequence identity between the mutant proteins and the mesophilic sequence is 85% and 78%, respectively. The mutant proteins differ from the thermophilic (73% identity for mutant A and 67% identity for mutant B) and hyperthermophilic (64% identity for mutant A and 59% identity for mutant B) variants as well as each other (93% identity). Mutant A shows a dielectric response ($\varepsilon_p = 16.4$) similar to *Bc*-Csp, and mutant B ($\varepsilon_p = 29.0$) is similar to the hyperthermophile *Tm*-Csp at 298K. The electrostatic contribution to stability was calculated and compared with the naturally occurring sequences from this 'thermophilic series' (Fig. 5). In general, we see an increased stability of the mutants over the mesophilic protein and similar behavior for mutant proteins and the corresponding (hyper-)thermophile; i.e., a similar dielectric constant at room temperature is observed for *Bc*-Csp with mutant A and *Tm*-Csp with mutant B. Experimental data regarding these mutants show smaller increases in the melting temperature relative to the wild-type mesophilic protein than anticipated from our electrostatic modeling studies. However, further studies have demonstrated that the electrostatic properties of these mutants are remarkably similar to the thermophilic species.[22] We believe greater thermostability could be achieved by minimizing perturbations to hydrophobic components of protein stability, which were not explicitly included in the current study. Our results suggest that the principles of design developed from the analysis presented here provide a reasonable framework for exploration of the origins of thermal stability in proteins, and that these principles may be used to guide the experimental development of proteins with enhanced stability to both extremes in cold and heat.[22,28]

CONCLUSIONS

We have demonstrated that the internal dielectric constant of a protein is correlated with thermostability in the cold shock and CheY protein families. We have argued that the dielectric response generally increases with the number of charges on the protein surface. Since hyperthermophiles display a greater number of charges relative to their mesophilic homologues, we propose that this class of

proteins will also exhibit a greater dielectric response. The increased dielectric response, in addition to the greater number of charge pairs, may stabilize this class of proteins in a global manner, consistent with experimental observations. This stabilizing effect has been demonstrated for both a simple two dielectric continuum model and through consideration of more complex, multi-dielectric Poisson models. Additionally, this theory suggests that the internal dielectric response of proteins can lead to cold denaturation by further reducing the solvating ability of the proteins arguing for a significant electrostatic contribution to cold denaturation, consistent with past work. Generally, our findings suggest that a more thoughtful consideration of the internal dielectric response of proteins may be important for understanding the role of electrostatic properties in biological systems from stability to ligand binding to macromolecular association, and that established ideas based on theoretical models assuming modest protein dielectric responses ($\varepsilon_p = 2–4$) may need to be re-examined in light of our findings.[73]

ACKNOWLEDGMENTS

The authors acknowledge financial support from the National Institutes of Health through grant GM37554.

REFERENCES

1. Madigan MT, Marrs BL. Extremophiles. Sci Am 1997;276:82.
2. Jaenicke R, Böhm G. The stability of proteins in extreme environments. Curr Opin Struct Biol 1998; 8:738–48.
3. Vogt G, Argos P, Protein thermal stability: hydrogen bonds or internal packing? Fold Des, 1997; 2:S40–S6.
4. Vogt G, Woell S, Argos P. Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs J Mol Biol, 1997; 269:631–43.
5. Karshikoff A, Ladenstein R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads Trends Biochem Sci, 2001; **26:**550–6.
6. Rees DC, Adams MWW. Hyperthermophiles: taking the heat and loving it Structure, 1995; 3:251–4.
7. Jaenicke R. Protein stability and molecular adaptation to extreme conditions Eur J Biochem, 1991; 202:715–28.
8. Querol E, Perez-Pons JA, Mozo-Villarias A. Analysis of protein conformational characteristics related to thermostability Prot Eng, 1996; 9:265–71.
9. Hollien J, Marqusee S. A Thermodynamic Comparison of Mesophilic and Thermophilic Ribonuclease H Biochemistry, 1999; 38:3831–6.
10. Hollien J, Marqusee S. Structural Distribution of Stability in a Thermophilic Enzyme Proc Natl Acad Sci, USA, 1999; 96: 13674–8.
11. Mayo SL, Malakauskas SM. Design, Structure and Stability of a Hyperthermophilic Protein Variant Nature Struct Biol, 1998; 5:470–5.
12. Lebbink JHG, Knapp S, van der Oost J, Rice D, Ladenstein R, de Vos WM. Engineering Activity and Stability of Thermotoga maritima Glutamate Dehydrogenase. II: Construction of a 16-Residue Ion-pair Network at the Subunit Interface J Mol Biol, 1999; 289:357–69.
13. Loladze VV, Ibarra-Molero B, Sanchez-Ruiz JM, Makhatadze GI. Engineering a thermostable protein via optimisation of charge-charge interactions on the protein surface Biochemistry, 1999; 38:16419–164223.
14. Jiang X, Bishop EJ, Farid RS. A *De Novo* designed protein with properties that characterize natural hyperthermophilic proteins J Am Chem Soc, 1997;119:838–9.
15. Matthews BW, Nicholson H, Becktel WJ. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding Proc Natl Acad Sci USA, 1987; 84:6663–7.
16. Macedo-Ribeiro S, Darimont B, Sterner R, Huber R. Small Structural Changes Account for the High Thermostability of 1[4Fe=4S] ferredoxin from the hyperthermophilic bacterium Thermotoga maritima Structure, 1996; 4:1291–301.
17. Thompson MJ, Eisenberg D. Transproteomic Evidence of a Loop-Deletion Mechanism for Enhancing Protein Thermostability J Mol Biol, 1999; 290:595–604.
18. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV. The complete genome of the hyperthermophilic bacterium Aquifex Aeolicus Nature, 1988; 392:353–8.
19. Rajdeep D, Gerstein M. The Stability of Thermophilic Proteins: A Study Based on Comprehensive Genome Comparison Fuct Integr Genomics, 2000;1:76–88.
20. Hashimoto H, Inoue T, Nishioka M, Fujiwara S, Takagi M, Imanaka T, Kai Y. Hyperthermostable Protein Structure Maintained by Intra and Inter-helix Ion-pairs in Archaeal O6-Methylguanine-DNA Methyltransferase J Mol Biol, 1999; 292:707–16.
21. Isupov MN, Fleming TM, Dalby AR, Crowhurst GS, Bourne PC, Littlechild JA. Crystal Structure of the Glyceraldehyde-3-phosphate Dehydrogenase from the Hyperthermophilic Archaeon Sulfolobus solfataricus J Mol Biol, 1999; 291:651–60.
22. Dominy BN, Perl D, Schmid FX, Brooks CL, III. The Effects of Ionic Strength on Protein Stability: The Cold Shock Protein Family J Mol Biol, 2002; 319:541–54.
23. Vetriani C, Maeder DL, Tolliday N, Yip KS-P, Stillman TJ, Britton K L, Rice DW, Klump HH, Robb FT. Protein thermostability above 100°C: A key role for ionic interactions Proc Natl Acad Sci USA, 1998; 95:12300–5.
24. Zhou HX, Dong F. Electrostatic contributions to the stability of a thermophilic cold shock protein Biophys J, 2003; 84:2216–22.
25. Xiao L, Honig B. Electrostatic Contributions to the Stability of Hyperthermophilic Proteins J Mol Biol, 1999; 289:1435–44.
26. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis Prot Sci, 1994; 3:211–26.
27. Honig B, Yang A. Free energy balance in protein folding Anfinsen CB, Edsall JT, Eisenberg DS, Richards. FM, editors. In: Advances in Protein Chemistry, Vol. 46. San Diego: Academic Press; 1995; p 27–58.
28. Perl D, Mueller U, Heinemann U, Schmid FX. Two Exposed Amino Acid Residues Confer Thermostability on a Cold Shock Protein Nature Struct Biol, 2000; 7:380–3. 380–3.
29. Elcock AH. The Stability of Salt Bridges at High Temperatures: Implications for Hyperthermophilic Proteins J Mol Biol, 1998; 284:489–502.
30. Grimsley GR, Shaw KL, Lanette RF, Alston RW. Huyghues-Despointes BMP, Thurlkill RL, Scholtz JM, Pace CN. Increasing protein stability by altering long-range coulombic interactions Prot Sci, 1999; 8:1843–9.
31. Perutz MF, Raidt H. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2 Nature, 1975; 255: 256–9.
32. Usher KC, de la Cruz AFA, Dahlquist FW, Swanson RV, Simon MI, Remington SJ. Crystal structures of CheY from *Thermotoga maritima* do not support conventional explanations for the structural basis of enhanced thermostability Prot Sci, 1998; 7:403–12.
33. Willimsky G, Bang H, Fisher G, Marahiel MA. Characterization of *cspB,* a *Bacillus subtilis* Inducible Cold Shock Gene Affecting Cell Viability at Low Temperatures J Bacteriol, 1992; 174:6326–35.
34. Schröder K, Zuber P, Willimsky G, Wagner B, Marahiel MA. Mapping of the *Bacillus subtilis* cspB gene and cloning of its homologs in the thermophilic, mesophilic and psychrotrophic bacilli Gene, 1993; 136:277–80.
35. Welker C, Böhm G, Schurig H, Jaenicke R. Cloning, overexpression, purification, physicochemical characterization of a cold shock protein homolog from the hyperthermophilic bacterium Thermotoga maritima Prot Sci, 1999; 8:394–403.
36. Bischoff DS, Ordal GW. Sequence and Characterization of *Bacillus subtilis* CheB, a Homolog of *Escherichia coli* CheY, and Its Role in a Different Mechanism of Chemotaxis J Biol Chem, 1991; 266:12301–5.
37. Schindelin H, Marahiel MA, Heinemann U. Universal nucleic acid- binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein Nature, 1993; 364:164–8.
38. Schnuchel A, Wiltscheck R, Czisch M, Herrler M, Willimsky G, Graumann P, Marahiel MA, Holak TA. Structure in solution of the

major cold- shock protein from Bacillus subtilis Nature, 1993; 364:169–71.
39. Tramontano A, Leplae R, Morea V. Analysis and assessment of comparative modeling predictions in CASP4 Proteins, 2001; Suppl 5:22–38.
40. Sanchez R, Sali A. Large-Scale Protein Structure Modeling of the Saccharomyces cerevisiae genome Proc Natl Acad Sci, USA, 1998; 95:13597–602.
41. Mueller U, Perl D, Schmid FX, Heineman U. Thermal stability and atomic-resolution crystal structure of the Bacillus caldolyticus cold shock protein J Mol Biol, 2000; 297:975–88.
42. Jorgensen WL, Chandrasekhar J, Madura J, Impley RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water J Chem Phys, 1983; 79:926–35. 926–35.
43. MacKerell J, AD, Bashford D, Bellott M, Dunbrack J, RL, Evanseck J D, Field MJ, Fischer S, Gao H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher I, WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz–Kuczera J, Yin D, Karplus M. All-atom Empirical Energy Function for Molecular Modeling and Dynamics Studies of Proteins J Phys Chem B, 1998; 102:3586–616.
44. Dominy BN, Brooks CL, III. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids J Phys Chem B, 1999; 103:3765–73.
45. Reiher WE, III. Theoretical studies of hydrogen bonding PhD thesis. Harvard University, 1985.
46. Neria E, Fisher S, Karplus M. Simulation of activation free energies in molecular systems J Chem Phys, 1996; 105:1902–21.
47. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, and Karplus M. CHARMM: A Program for Macromolecular Energy, Minimization, and Molecular Dynamics J Comp Chem, 1983; 4:187–217.
48. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*- Alkanes J Comp Phys, 1977; 23:327–41.
49. Simonson T, Brooks CL, III. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics J Am Chem Soc, 1996; 118:8452–8.
50. Sham YY, Muegge I, Warshel A, The Effect of Protein Relaxation on Charge-Charge Interactions and the Dielectric Constant of Proteins Biophys, J, 1998; 74:1744–53.
51. Sham YY, Chu ZT, Warshel A. Consistent Calculations of pK_a's of Ionizable Residues in Proteins: Semi-microscopic and Macroscopic Approaches J Phys Chem B, 1997; 101:4458–72.
52. Zhou Y, Vitkup D, Karplus M. Native Proteins are Surface-molten Solids: Application of the Lindemann Criterion for the Solid *versus* Liquid State J Mol Biol, 1999; 285: 1371–5.
53. Qiu D, PS, S, Hollinger FP, Still WC. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii J Phys Chem A, 1997; 101:3005–14.
54. Warwicker J, Watson HC. Calculation of the Electric Potential in the Active Site Cleft due to α-Helix Dipoles J Mol Biol, 1982; 157:671–9.
55. Gilson MK, Honig B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis Proteins, 1988; 4:7–18.
56. Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free- Energies using Macroscopic Solvent Models J Phys Chem, 1994; 98:1978–88.
57. Gilson M, Sharp K, Honig B. Calculating the Electrostatic Potential of Molecules in Solution - Method and Error Assessment J Comp Chem, 1988; 9:327–35.
58. Jean-Charles A, Nicholls A, Sharp K, Honig B, Tempczyk A, Hendrickson TF, Still WC. Electrostatic Contributions to Solvation Energies: Comparison of Free Energy Perturbation and Continuum Calculations J Am Chem Soc, 1991; 113:1454–5.
59. Strop P, Mayo SL. Contribution of Surface Salt Bridges to Protein Stability Biochemistry, 2000; 39:1251–5.
60. Waldburger CD, Schildbach JF, Sauer RT. Are Buried Salt Bridges Important for Protein Stability and Conformational Specificity? Nature Struct Biol, 1995; 2:122–8. 122–8.
61. Wimley WC, Gawrisch K, Creamer TP, White SH. Direct Measurement of Salt-Bridge Solvation Energies using a Peptide Model System: Implications for Protein Stability Proc Natl Acad Sci, USA, 1996; 93:2985–90.
62. Aguilar CF, Sanderson I, Moracci M, Ciaramella M, Nucci R, Rossi M, and Pearl LH. Crystal Structure of the b-Glycosidase from the Hyperthermophilic Archeon Sulfolobus solfataricus: Resilience as a Key Factor in Thermostability J Mol Biol, 1997; 271:789–802.
63. Mozo-Villarias A, Cedano J, Querol E. A simple electrostatic criterion for predicting the thermal stability of proteins Prot Eng, 2003; 16:279–86.
64. Brooks CL, III, Karplus M, Pettitt BM. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics, Vol. 71. New York: Wiley; 1988.
65. Archer DG, Wang P. The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes J Phys Chem Ref Data, 1990; 19:371.
66. Li W-T, Grayling RA, Sandman KSE, Shriver JW, Reeves JN. Thermodynamic Stability of Archaeal Histones Biochemistry, 1998; 37:10563–72.
67. Grättinger M, Dankesreiter A, Schurig H, Jaenicke R. Recombinant Phosphoglycerate Kinase from the Hyperthermophilic Bacterium Thermotoga maritima: Catalytic, Spectral and Thermodynamic Properties. J Mol Biol, 1998; 280:525–33.
68. Loncharich RJ, Brooks BR. Temperature of Force Field Calculations with Neutron Scattering Data J Mol Biol, 1990; 215:439–55.
69. Rasmussen BF, Stock AM, Ringe D, Petsko GA. Crystalline ribonuclease A loses function below the dynamical transition at 220K Nature, 1992, 357:423–4.
70. Brooks BR, Dusnaka J, Karplus M. Harmonic Analysis of Large Systems. I. Methodology J Comp Chem, 1995; 16:1522–42.
71. Graziano G, Catanzano F, Riccio A, Barone G. A Reassessment of the Molecular Origin of Cold Denaturation J Biochem, 1997; 122:395–401.
72. Makhatadze GI, Privalov PL. Energetics of protein structure In "Adv Protein Chem", Anfinsen CB, Edsall JT, Richards FM, Eisenberg DS, eds. Vol. 47. San Diego: Academic Press; 1995; p 307–425.
73. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins, 2001; 44:400–17.
74. Szilagyi A, Zavodszky P. Structural Differences between Mesophilic, Moderately Thermophilic, and Extremely Thermophilic Protein Subunits: Results of a Comprehensive Survey Structure, 2000; 8:493–504.

APPENDIX

General Features of Protein Dielectric Properties

In order to demonstrate a general connection between the number of surface charges and dielectric response, a simplified protein model was created. The model was derived by starting with the structure of Buckminster fullerene, a spherical molecule comprised of 60 evenly spaced carbon atoms. A radius of gyration (R_g) restraint was applied to the structure, and the structure was minimized to satisfy an R_g value of 12.9 Å, resembling a small protein. Dynamic motion was imparted by eliminating the bonds in the structure and applying a harmonic restraint to each carbon atom about its position on this sphere of ~13 Å. The force constant of the harmonic restraint was consistent with root mean square fluctuations of ~2.6 Å at 298K (k = 0.6923 kcal mol^{-1} Å^{-2}). The value of 2.6 Å was derived from the maximum observed fluctuation of a charged residue from the implicit solvent molecular dynamics simulation of *Bs*-CspB. A MD simulation was then performed on the simplified protein model for 2 ns, where each of the 60 atoms/particles moved independently in a harmonic potential well. Using this structural and dynamic model, balanced charges (+1, −1) of varying number were assigned to random carbon atom pairs. This maintained a net zero charge regardless of the total number of charges. A dielectric constant was then computed for various charge distributions and identical

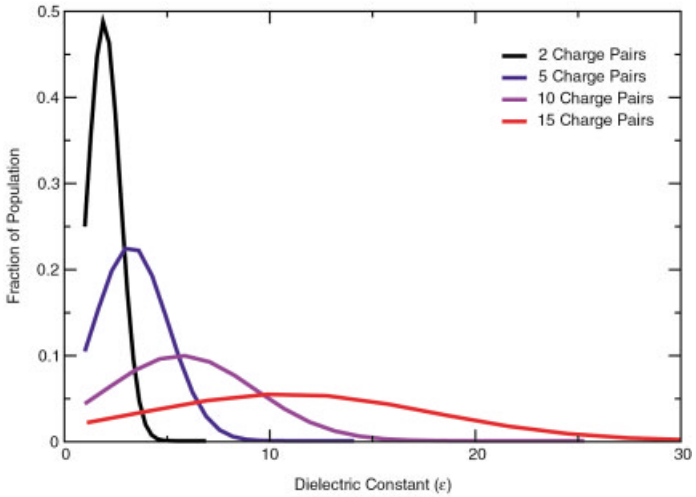

Fig. A1. The distributions of dielectric response for random charge distributions of a fixed number of charges assigned to a simplified protein model. For a fixed number of charges, multiple distributions of those charges and therefore dielectric constants are possible. The results are the Gaussian distributions of dielectric response shown in the figure.

dynamic motions, using the same method applied to the atomic resolution of cold shock and CheY proteins. The result, shown in Figure A1, was a Gaussian distribution of dielectric constant values for each fixed number of charges. As the number of charges was increased, the mean in the distribution was shifted to higher dielectric values. This indicates that, as more charges are added, even to this simple protein model, the dielectric constant is expected to increase. Since we know from genetic and structural studies that the number of charges is found to be increased in thermophilic and particularly hyperthermophilic proteins relative to their mesophilic homologues,[18,74] we also expect their dielectric response to be higher, as seen in the Csp and CheY families.

Although the mean dielectric response (most probable dielectric response) does increase as the charge density increases, the model also accounts for exceptions to this trend. The distributions of dielectric constants have significant variance and overlap with each other, indicating that systems with fewer charges may exhibit a greater dielectric response than systems with greater numbers of charges. The placement of charges, and not solely the number of charges, can play a significant role in determining the dielectric response of a system. This phenomenon is also observed in the Csp family. The thermophilic protein *Bc*-Csp has fewer charged residues than the mesophilic *Bs*-CspB, yet *Bc*-Csp also exhibits a larger dielectric response.