Turning a Mesophilic Protein into a Thermophilic One: A Computational Approach Based on 3D Structural Features

Sohini Basu and Srikanta Sen*

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

Received February 25, 2009

In spite of considerable improvement of our understanding of factors responsible for protein thermostability, rational designing of thermostable variants of mesophilic proteins is not yet fully established. The present paper describes an effective computational strategy that we have developed to identify the most suitable mutations converting a chosen mesophilic protein into a thermophilic one starting from its 3D structure. The approach is based on the concept that stabilization of several surface residues should enhance the global stability of the protein. The method relies on the estimation of electrostatic and van der Waals interactions in computing the interaction among the side chains of individual residues and the rest of the protein. The polar or charged residues whose side chains interact weakly with the rest of the original protein are identified first. Then, for each such identified residue (A), another residue (B) in its spatial vicinity is identified. The side chain of the residue (B) is then replaced by a suitable conformer of a residue that is electrostatically complementary to the residue (A) to enhance local interactions and hence the stability of the protein. The steric effect is taken care of through van der Waals interactions. We reject the mutations that improve interactions only locally along the sequence as it is unlikely to enhance the global stability of the 3D architecture. We use the difference in self-energies (ΔE_{self}) as a measure of the stability difference between the original and its mutant variant. This paper presents two test cases with demonstration of the enhanced stability of such mutated proteins and validates the strategy by considering five experimentally known thermophilic−mesophilic protein pairs.

INTRODUCTION

Over the past decade, a large number of microorganisms have been discovered that are natural habitants of extreme conditions like alkaline lakes, volcanic systems, and extremely acidic environments, etc. They are commonly called extremophiles and include extreme varieties as thermophiles, halophiles, acidophiles, alkalophiles, psychrophiles, osmophiles, barophiles, and radiation resistant organisms. The organisms that are found in hot places like hot springs and hot vents at the bottom of the sea are known as thermophilic organisms. Depending on the optimal growth temperature (T_{opt}) thermophilic organisms are generally divided into two classes (i) moderate thermophiles (50 °C < T_{opt} < 80 °C) and (ii) hyperthermophiles (T_{opt} > 80 °C). It is of fundamental interest to understand the molecular basis of how thermophilic proteins gain and maintain their stability and activities at high temperatures in spite of the fact that such proteins are also made of the same building blocks as the mesophilic ones are. Numerous works on the physicochemical origin of thermostability have been made, and it has been suggested that a variety of factors can contribute to this observed enhanced stability.[1−11] Such factors include hydrophobic interactions, improved packing, networks of hydrogen bonds, optimized electrostatic interactions,[5,6,12−17] core packing, and cavity filling etc.[18,19] It is now believed that the enhancement of stability in thermophilic proteins results from a number of locally improved interactions in 3D.

Recent works have indicated that hyperthermophiles are generally characterized by an increased number of surface charges and ionic networks that play dominant roles in enhancing thermostability.[7,20,21] It was further demonstrated that the desolvation cost for the formation of a salt bridge decreases at elevated temperatures, leading to an effective stabilization.[5,22] Thus, it appears that ion pair interactions may have considerable roles in enhancing the overall stability of thermophiles. The importance of the dynamic arrangement of ion pairs and their individual contributions to the overall thermal stability of a protein has recently been demonstrated by MD simulation of the crystal structures of thermophilic and mesophilic proteins by Danciulescu et al.[23] Moreover, in proteins there are two types of H-bond forming residues, (i) charge neutral and (ii) with a charge, and thus there may be H-bonds between charged-neutral and neutral−neutral residues. It has been suggested[24] that the desolvation energy for making a H-bond is lower than that for an ion pair and the binding energy of a charged-neutral H-bond is far larger than that from neutral−neutral H-bonds, due to the charge-dipole interaction. These imply that charged residues might have an additional influence on stability through the formation of H-bonds. Hydrophobic interactions prevent unfavorable aqueous solvation and also improve van der Waals interactions by cavity filling. Hydrophobic interaction is another factor that is known to stabilize proteins to a large

* Corresponding author phone: +91 33 23572744/45; fax: +91 33 23574113; e-mail: srikanta@chembiotek.com.

10.1021/ci900183m CCC: $40.75
Published on Web 07/08/2009

extent at physiological temperature, but its role in thermophilic proteins is not yet clearly understood.

Due to the immense potential for the important biotechnological applications of proteins that are capable of functioning at higher temperatures, there is much interest in thermophiles. Mainly for this reason, it is highly desirable to develop methods of directed engineering toward higher thermostability for chosen normal mesophilic proteins. Successful design of thermophilic protein also indicates the level of our understating of the mechanisms of thermostability. There are a few approaches for making a given protein a thermophilic one.[25−30] Dantas et al. reported that based on their computational strategies they successfully designed nine different proteins, and a dramatic increase in the stabilities of these proteins was experimentally demonstrated.[30] However, these methods depend heavily on the sequence comparisons, and it may happen that a specific amino acid bias in thermostable proteins is more related to the evolutionary changes than a direct relevance to its thermostability. These methods also do not take into account the 3D features of the proteins.

In the present work we have utilized our current knowledge to develop a computational method to convert a mesophilic protein into a thermophilic one through several key mutations. Keeping consistent with the current beliefs that charge interactions are the most important factors in enhancing the stability of a thermophile we increase their number on the surface of the selected protein without any inter-residue steric conflicts. The main issue of the method is to identify the polar or charged residues that are involved in low energies of interaction with the rest of the protein. Then, in turn, for each such residue, another residue in its vicinity is identified and replaced by a matching conformer of a suitable residue in order to enhance its interaction with the neighboring residues of the protein and thus stabilize the protein further. We have presented two case studies to demonstrate the capability of the present method and validated the methods by considering five known thermophilic−mesophilic protein pairs including a very well-studied protein pair. The results indicate that the present approach is an easy and efficient method that can be successfully employed to convert any protein into a thermophilic one. The present designing method differs from the other methods in three major ways that (i) we do not use sequence comparison for finding the desired mutations, (ii) protein's 3D features and inter-residue physical interactions are explicitly considered, and (iii) desirable mutations are identified on the basis of interaction energies to stabilize tertiary contacts. We believe that this is the first report suggesting computational strategy in converting a mesophilic protein into a thermophilic one using only its 3D structure.

METHODS

It is quite likely that the thermal denaturation of a protein starts at the protein surface. Openings are created at the surface due to thermal fluctuation, solvent molecules enter into the protein body, more denaturation of the inside part of the protein occurs, and the denaturation process continues. If the surface of a protein is made resistant to large fluctuations, the denaturation of the whole protein at a given temperature may be prevented. Thus, our basic strategy is to mutate a set of preselected less stable surface residues to achieve enhanced stability. There are also publications[31−36] that demonstrate that residues at the surface do play important roles in enhancing the thermostability of thermophilic proteins. Moreover, we concentrate only on the surface residues as changes in them are not likely to alter the 3D architecture of the protein. Conversely, due to the tight packing changing the residues in the interior of a protein may alter the 3D architecture and hence its functional activity.

Our strategy consists of a set of defined steps.

(i) First we look into the residue-wise interaction pattern of the original energy-minimized 3D structure of the protein in order to identify the polar or charged residues whose side chains interact weakly with the rest of the protein and are considered as the key residues. This interaction energy includes electrostatic and van der Waals components.

(ii) Then we consider only such key residues that are close to the surface, and subsequently we identify some other nearby residues that interact weakly with the selected key residue. The side chains of such residues are chosen for replacement by desired residues in order to improve its interaction with the key residue and hence the stability of the protein.

(iii) The side chains of each of these selected residues are then replaced by a conformer of the side chain of a residue that produces the best improvement of the interaction energy of it with the rest of the protein. This process is repeated for all the selected residues corresponding to all the selected key residues.

(iv) Finally, the resulting protein is energy minimized by CHARMM considering the full force field.[37,38]

(v) The self-energies of the original protein and the mutated protein are computed and compared to demonstrate the substantial gain in stability of the mutated protein. Self-energy (E_{self}) of a protein in its 3D structure represents the total energy content of the structure. The stability of a protein is directly related to the difference between its folded and the unfolded forms. However, as in the present case we are only interested in the difference in stability, we use the difference in self-energies (ΔE_{self}) as a measure of the stability difference between the folded structures of the original and its designed mutant variant.

The entire process can be repeated for a few cycles to make the effect more pronounced if required. The detailed methods of each step are given below.

Preparing the Side Chains Conformer Libraries. In order to mutate a residue, we need conformer libraries of the side chains for all the residues. In generating the side-chain conformer libraries we randomly downloaded a number of crystal structures of proteins from the RCSB database. From that lot we finally selected only 49 PDBs that were single chain globular structures without any missing residues in the respective PDB files. A list of the PDB IDs of these 49 proteins is available in the Supporting Information.

First, we downloaded randomly chosen 49 pdb files of proteins. Then, H atoms were assigned to each of these pdb files. Each H-assigned pdb file was energy minimized in vacuum by 5000 Steepest Descent steps using CHARMM energy minimization tool and parameters[37,38] in order to remove bad steric conflicts from the structure. In the energy minimization process we have used a distance dependent

dielectric constant with an epsilon value of 1.0 and have used a spherical cutoff of 12.0 Å in computing the nonbonded interactions. All 49 chosen protein structures (pdb files) were processed in this way. Subsequently, we extracted the coordinates of the side chains of a particular residue type and stored it in a specific file named by its type such as Lys.lib. This way we generated the conformer libraries for each of the 20 different residue types. The advantage of generating the side-chain conformer library directly from the crystal structures is that it provides the side-chain conformations that are truly allowed in nature. We have developed in-house *FORTRAN* codes for generating the conformer libraries from the pdb files of the proteins.

Identifying Weakly Interacting Polar/Charged Side Chains and Their Mutatable Partner Residues. Considering electrostatic and *van der Waals* interactions the energies of interaction of the individual side chains with the rest of the protein are computed using in-house *FORTRAN* codes. The electrostatic interaction was computed by *Coulomb's* law using the standard relation $E_{elec} = 331.5 \times \sum_{i \neq j} \sum_{j=1}^{n} (q_i q_j)/(\varepsilon . d_{ij})$ where q_i represents the partial atomic charge of the i^{th} atom of the respective side chain (excluding the backbone part) of the residue of interest, and q_j represents the same for the atoms of the rest of the protein. We used a distance cutoff value of 12.0 Å and a distance dependent dielectric constant (ε) as in CHARMM. The *van der Waal* interaction was computed using the *Lennard-Jones* formula $E_{vdw} = \sum_{i \neq j} \sum_{i=1}^{n} \varepsilon_{ij} [((\sigma_{ij})/(d_{ij}))^{12} - 2((\sigma_{ij})/(d_{ij}))^{6}]$, where $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$, where σ_i and σ_j represent the *Lennard-Jones* diameters of the i^{th} and j^{th} atoms respectively, d_{ij} corresponds to the interatomic distance between the i^{th} and j^{th} atoms, and $\varepsilon_{ij} = (\varepsilon_i \varepsilon_j)^{1/2}$, with ε_i and ε_j being the *Lennard-Jones* well-depth of the i^{th} and j^{th} atoms. We used the partial atomic charges and the *van der Waal* parameters following CHARMM. It should be pointed out here that we used our in-house tool only to identify the key residues and the most suitable mutation in its neighborhood. The final energetics in all cases were computed using CHARMM.

Using the generated residue-wise interaction pattern, the polar or charged residues that are interacting weakly with the rest of the protein are identified and are considered as the key residues. These weak interactions may be a result of the fact that the residue is placed in a noncomplementary chemical environment. In such cases, a residue in its immediate vicinity may be identified and subsequently replaced by a polar or charged residue of complementary nature to gain in stability. It should be emphasized that all these factors depend on the 3D structure of the protein, and contributions come from interactions among many residues that are 'sequence-wise local' as well as 'nonlocal along the sequence'. Only 'sequence-wise nonlocal' interactions are expected to be the major determinants of the stabilization of the 3D structure of a protein.

Based on this idea, we select several (10−15) such key polar or charged side chains distributed closely over the protein surface. Subsequently, we identify potential residues for side-chain substitutions, by visually inspecting the immediate neighborhood of the key residues in the energy-minimized protein structure. Then, as described below, we replace the individual side chains by the best conformer of each of the other 19 residues obtained by searching the conformer libraries in order to find the best substitution.

If the key residue and its mutatable partner are very close (less than 10 residues) to each other along the sequence, we do not consider that as it is quite unlikely that this will effectively enhance the overall stability of the protein. This criterion eliminates the possibility of false selections due to favorable interactions among 'sequence-wise local' residues as those favorable interactions are expected to be present also in the unfolded forms of the protein and hence should be ineffective in the stability of the folded protein.

Furthermore, in our strategy, the mutations are made only to the residues on or very close to the surface of the protein. Hence, it is highly unlikely that due to such mutations the 3D architecture of the protein will be affected significantly. Thus, the proteins designed following our strategy are highly expected to fold and function in a proper way.

Finding the Suitable Substitutes and Generation of the 3D Structure of the Designed Protein. We have developed in-house *FORTRAN* codes to identify the best substituting residue side chains by trying all the conformers of the different side chains from the conformer libraries generated earlier. For each type of residue, all the individual side-chain conformers are substituted, and its interaction energy with the rest of the protein is computed as explained before and the conformer giving the lowest energy of interaction with the rest of the protein is identified. This is repeated for all 19 residues. We select the best replacement considering the most favorable interaction with the rest of the protein and replace the original side chain by it. These steps are repeated to replace all the selected mutatable side chains one by one. It may be pointed out here that in each case we also identify the residues that contributed mostly to the interaction with the selected conformer. At the end of the process we get the atomic coordinates of the designed protein with the substituted side chains.

The generated structure of the new mutated protein is then energy minimized, and the self-energy of the protein is computed again using CHARMM. The self-energy of the designed protein is expected to be significantly lower than that of the original protein indicating enhanced stability of the designed protein. It may be emphasized that as the protein has changed the difference in self-energy should not strictly represent the difference in stability but is a reasonable indicator of it.

Demonstration and Validation of Enhanced Stability. The ideal way to demonstrate the enhancement of stability of the designed protein could be to synthesize the designed mutated protein and compare the thermal stabilities by performing melting experiments. However, it is beyond our scope. As an alternative, we can use computational measures for this purpose. Comparison between the self-energies of the energy minimized original and the mutated proteins is a good way of demonstrating the differences in their thermal stabilities. Thus, the difference ($\Delta E_{self} = E_{self}^{Th} - E_{self}^{Me}$) in the self-energies of the thermophilic (E_{self}^{Th}) and mesophilic version (E_{self}^{Me}) can be used as a measure of the stability differences between the two. As defined here, a negative value of ΔE_{self} implies that the thermophilic variant is more stable than the corresponding mesophilic one. The residue-wise plot of the differences in the interaction energies of the respective individual residues of the protein pair also represents the way the change in stability is affected at the

Table 1. Summary of the Selected Key Residues, Identified Mutatable Residues in Its Vicinity, the Best Mutation, and the Respective Computed Energy Differences[a]

PDB id	sequence identity	no.	key residue	mutatable residue	Δres	mutation made	ΔE_i (kcal/mol)	D_{CA} (Å)
		1	Lys-21	Gln-124	103	Asp-124	−27.4	8.4
		2	Asp-43	Asn-5	38	Lys-5	−16.1	4.9
		3	Asp-47	Trp-88	41	Arg-88	−41.6	10.3
2rcq	95%	4	Arg-58	Thr-29	29	Asp-29	−23.5	9.5
		5	Glu-69	Leu-85	16	Arg-85	−46.9	4.8
		6	Glu-72	Val-62	10	Lys-62	−54.4	7.8
		7	Asp-123	Ala-22	101	Lys-22	−42.5	7.6
		1	Glu-33	Phe-152	119	Lys-152	−53.6	7.8
		2	Gln-42	Val-10	32	Lys-10	−52.7	7.9
		3	Arg-75	Phe-160	85	Asp-160	−28.3	13.4
2bl1	95.9%	4	Gln-102	Ile-5	97	Arg-5	−63.2	7.7
		5	Arg-112	Leu-163	51	Asp-163	−13.7	10.9
		6	Arg-136	Val-101	35	Glu-101	−36.1	9.6
		7	Asp-157	Pro-146	11	Lys-146	−62.9	4.9

[a] $\Delta E_i = E_i^{Th} - E_i^{Me}$ represent the difference in the energies E_i^{Th} and E_i^{Me} of interactions of the i^{th} residue with the rest of the thermophilic variant and the mesophilic protein respectively, for the two test cases. Δres represents the number of residues separating the key residue and the respective mutatable residue along the sequence. D_{CA} represents the spatial distance (Å) between the C-alpha atoms of the key residue and the respective mutated residue. Sequence identity refers to the original protein and its mutated variant.

individual residue level. We have used these methods in some of our earlier works[39−42] as well.

It should also be pointed out that as the inter-residue interactions have been addressed at the atomic level, the applicability of the present method should not have any bias toward any secondary or tertiary structure.

In order to validate our method of demonstrating stability differences between proteins, we have applied it to five widely different known mesophilic−thermophilic protein pairs including an experimentally very well-studied protein pair. The individual pdb files downloaded were first energy minimized using the same protocol as mentioned before, and ΔE_{self} values were computed for the protein pairs to compare the stabilities of the individual protein pair. The interaction energies of individual residues with the rest of the protein were also computed. We then aligned the two sequences of the thermophilic−mesophilic protein pair in order to identify the respective residues in the two proteins based on one to one mapping of the aligned residues. After that we removed the gaps by deleting the residues from the other protein and computed the ΔE_i values for the aligned respective residue pairs from the two protein sequences. The residue-wise interaction energy patterns were finally compared.

RESULTS

Conformer Libraries Generation. The conformer libraries for the different kinds of amino acid residues were generated by our in-house tool as explained in the Methods section. The conformers were visually checked and found to reasonably represent the conformational space available to the side chain. Depending on the size and flexibilities of the side chains, the numbers of conformers vary in the range of several hundreds to several thousands. However, the conformer libraries generated and used in the present work are reasonably large.

Selection of Test Systems. For simplicity's sake, we have selected two moderate sized single chain globular proteins (i) *human apo cellular retinol binding protein II* (*pdb id*: *2rcq*) and (ii) *phosphinothricin acetyltransferase* from *Pseudomonas aeruginosa* (*pdb id*: *2bl1*) with known experimental 3D structures, as representative mesophilic proteins to test our strategy of converting a mesophilic protein to a thermophilic one. The 3D crystallographic structures were collected from the RCSB database. H-atoms were assigned to each protein, and the structures were refined by energy minimization by 5000 steps of steepest descent algorithm with a distance dependent dielectric constant and a spherical cutoff of 12.0 Å in computing the nonbonded interactions employing the CHARMM molecular modeling package.[37,38]

Identifying Weakly Interacting Polar/Charged Side Chains and Their Mutatable Partner Residues. Examination of the residue-wise interaction pattern for the energy minimized protein 2rcq indicated that several polar/charged residues were associated with interaction energies much weaker compared to the same residue types at other positions. This implies that they do not have an electrostatically complementary residue in their immediate neighborhood. Out of these polar/charged residues, we selected Lys-21, Asp-43, Asp-47, Arg-58, Glu-72, Glu-69, and Asp-123 as the key residues whose interactions with the rest of the protein need to be enhanced. We visually inspected the neighborhood of each such key residue and identified a nearby residue that may be replaced by a residue of complementary nature. The key residues and the respective mutatable residues selected in this way are summarized in Table 1. The spatial distance (D_{CA}) of the CA atoms of the key-residue and its mutatable partner residue indicates that they are spatially close to each other particularly the side chains. It is further noticeable that none of the residue pairs (the key residue and its mutatable partner) are closer than 10 residues along the sequence. Thus, the mutations should enhance the overall stability of the 3D structure and not 'sequence-wise local' structure of the protein.

Similar calculations and analysis were also made for the other test case protein (*2bl1*), and the details are summarized in Table 1. In this case the residues Glu-33, Gln-42, Arg-75, Gln-102, Arg-112, Arg-136, and Asp-157 were selected as the key residues whose interactions with the rest of the protein need to be improved by mutations in their neighborhood. The respective mutatable residues selected are Phe-152, Val-10, Phe-160, Ile-5, Leu163, Val-101, and Pro-146. The Δres values in Table 1 further confirm that none of the key residues and its mutatable partner is 'sequence-wise

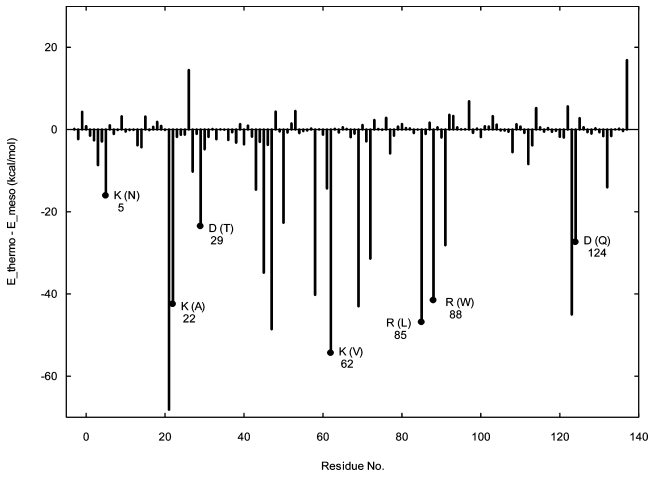

Figure 1. The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the test protein *2rcq* and its designed thermophilic variant is shown. The interaction energy of the ith residue of the test protein *2rcq* with the rest of the protein is represented by E_i^{Me}. Similarly, E_i^{Th} represents the same for the respective designed thermophilic variant. The residues selected for mutations are indicated by a filled circle (•) symbol at the end of the respective bar along with the mutated residue name indicated beside the bar with the original residue name within brackets. The residue numbers are also indicated explicitly.

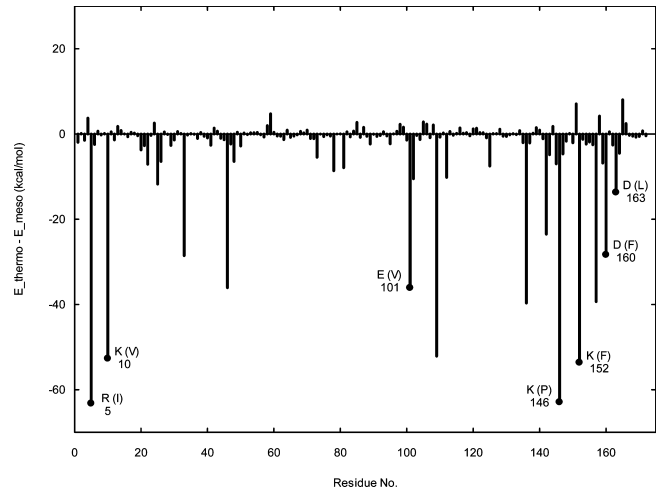

Figure 2. The differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the test protein *2bl1* and its designed thermophilic variant is shown. The interaction energy of the ith residue of the test protein *2bl1* with the rest of the protein is represented by E_i^{Me}. Similarly, E_i^{Th} represents the same for the respective designed thermophilic variant. The residues selected for mutations are indicated by a filled circle (•) symbol at the end of the respective bar along with the mutated residue name indicated beside the bar with the original residue name within brackets. The residue numbers are indicated explicitly.

close', and D_{CA} values ensure that they are 'spatially close'. Thus, the mutations should enhance the overall stability of the 3D structure.

Generating the Best Mutated Protein. For each selected mutatable residue in the test protein structure (2rcq) we replaced the respective side chain by the individual conformers of the side chains of the other residue types from the pregenerated conformer libraries and computed its interaction energy with the rest of the protein by using our in-house *FORTRAN* codes as mentioned in the method section. The atomic coordinates of the mutated side chains were directly generated by our in-house codes from the conformer libraries. For each mutatable residue, the atomic coordinates of the mutated side chain obtained in the previous step were substituted to the side chain of the respective mutatable residue. The mutated protein obtained in this way was then energy minimized by 5000 steps of steepest descent algorithm in order to remove any residual distortion or bad contacts. Table 1 summarizes the suggested substitution and the resulted gain in stability. All the best substituted side chains obtained are charged ones such as Asp, Lys, and Arg. It is clearly seen that substitution by each of these suggested side chains resulted in significant improvement in the interaction energy in the range (−16.1 to −54.4 kcal/mol) with the rest of the protein.

The same procedures were also carried out for the other test protein *2bl1*, and the results are summarized in Table 1. In this case also all the individual mutations improved the stability of the protein. The gain in interaction energy in the individual level is in the range (−13.7 to −62.9 kcal/mol). Here also the most suitable substituted side chains are found to be Asp, Glu, Lys, and Arg all of which have charged side chains.

Comparison of the Residue-Wise Interaction Patterns of the Original and Mutated Proteins and Demonstration of Stability Enhancement. Figure 1 compares the residue-wise interaction energy difference ($\Delta E_i = E_i^{Th} - E_i^{Me}$) profiles of the original protein *2rcq* and its designed mutated variant. The interaction energy of the ith residue with the rest of the thermophilic variant is represented as E_i^{Th}, while E_i^{Me} represents the same for the corresponding mesophilic variant of the protein pair. It is clearly seen that the residues that showed weaker interactions with the rest of the protein in the original protein are now showing stronger interactions in the corresponding mutated protein as expected. We also examined that as expected, the most favorably interacting partner of the mutated residues are found to be the respective key residues.

It is further interesting to note that not only the interaction energies of the mutated residues and the key residues are improved but also the interaction energies of several other residues are also improved significantly due to the mutations (Figure 1). Such residues are Asp-3, Phe-27, Asp-45, Lys-50, Asp-61, Asp-91, and Lys-132 whose interactions with the mutated residues increased the overall stability of the protein.

Figure 2 represents the same for the other protein *2bl1* selected as test system. Here also the growth of several strong and favorable interactions in the mutated protein is clearly demonstrated. It is found that all the 'mutated residues' have contributed significantly in gaining further stability. Only two residues Asp-160 and Asp-163 contributed relatively weakly to the gain in stability compared to the other mutated residues, while most effective contributions came from the residues Arg-5 and Lys-146. Apart from these mutated residues, several other residues like Gly-25, Asp-46, Glu-78, Arg-81, Glu-109, and Ser-142 participated significantly in stabilizing the protein in this changed scenario mostly due to their interactions with the mutated residues.

It should be clarified here that Figure 1 represents the differences (ΔE_i) in interaction energies of the individual residues in the mesophilic and thermophilic counterparts with the rest of the respective protein and do not strictly indicate that a residue is weakly interacting in mesophilic protein and strongly interacting in its designed thermophilic counterpart

Table 2. Comparison of the Self-Energies between the Mesophilic and Thermophilic Versions of the Two Test Proteins[a]

protein name	no. of residues	E_{self}^{meso} (kcal/mol)	E_{self}^{therm} (kcal/mol)	ΔE_{self} (kcal/mol)	ΔE_{self} *per residue* (kcal/mol)
2RCQ	141	−2096.3	−2482.3	−385.9	−2.74
2BL1	172	−2013.0	−2414.8	−401.8	−2.34

[a] $\Delta E_{self} = E_{self}^{th} - E_{self}^{me}$ (kcal/mol). The number of residues in the original protein and its mutated variant remains the same.

in absolute sense. However, the way ΔE_i values are defined and as the ΔE_i values are given with signs, it clearly indicates that large negative values of ΔE_i correspond to (relatively) stronger interactions in the thermophilic protein compared to that of mesophilic one. Moreover, as we are interested in the change in stabilities of the proteins ΔE_i appears to be a more relevant parameter than actual interaction energy values.

Table 2 summarizes the comparison of the differences in the overall self-energies (ΔE_{self}) between the thermophilic and mesophilic partners of the test proteins *2rcq* and *2bl1*. It is clearly demonstrated that in both test cases presented here, the designed themophilic variant is significantly more stable than the corresponding mesophilic ones even though the difference may not be drastic. The computed values of ΔE_{self} per residue further indicate that the degree of enhancement of stability for the protein *2rcq* (mutated) is more compared to that in the case of the protein *2bl1* (mutated). ΔE_{self} per residue provides a measure of gain in stability per residue and is a good measure to compare the stability gain between proteins of different sizes.

Stereoviews of the 3D structures of the energy minimized mutated variants for the test proteins with the key residues and mutated side chains highlighted are presented in Figures 3 and 4 for the two test systems. It is clearly seen that in both test cases, all the mutated residues are close to the surface of the protein and are well distributed over the protein's surface. Thus, a number of 'spatially local' stabilizations enhanced the global stability of the proteins.

Validation of the Strategy in Demonstrating Enhanced Stability. In order to validate this correlation between the differences in self-energies of the thermophilic−mesophilic protein pair to their relative stabilities, we considered five known thermophilic−mesophilic protein pairs (*1caa*, *8rxn*: *rubredoxin*), (*1yna*, *1xnb*: *zylanase*), (*1tmy*, *3chy*: *chey protein*), (*2prd*, *1ino*: *inorganic pyrophosphatase*), and (*1c9o*, *1csp*: *cold shock protein*). We energy minimized the H-atom assigned respective crystal structures as mentioned before and computed their self-energies. Table 3 summarizes the computed ΔE_{self} values with large negative values indicating a significant increase in the stability of the respective thermophilic variant compared to the mesophilic counterpart for four protein pairs. For the protein pair (*1tmy*, *3chy*) the computed value of ΔE_{self} was found to be +76.1 kcal/mol which is a consequence of the fact that the mesophilic partner (*3chy*) has a larger number of residues compared to that of the thermophilic counterpart (*1tmy*). As a result, *3chy* has a much lower value of E_{self} compared to that of *1tmy* causing a positive value of ΔE_{self}. It is interesting to note that the sequence identity between the proteins of this pair is very poor (29.1%). However, the stabilization effect is demonstrated by the negative value (−0.2 kcal/mol) of ΔE_{self} per residue even though it indicates that the stability difference of this protein pair may be much lesser compared to the other protein pairs considered here.

We further examined the ΔE_i vs i plots (Figure 5a-e) for the five protein pairs and observed similar features as

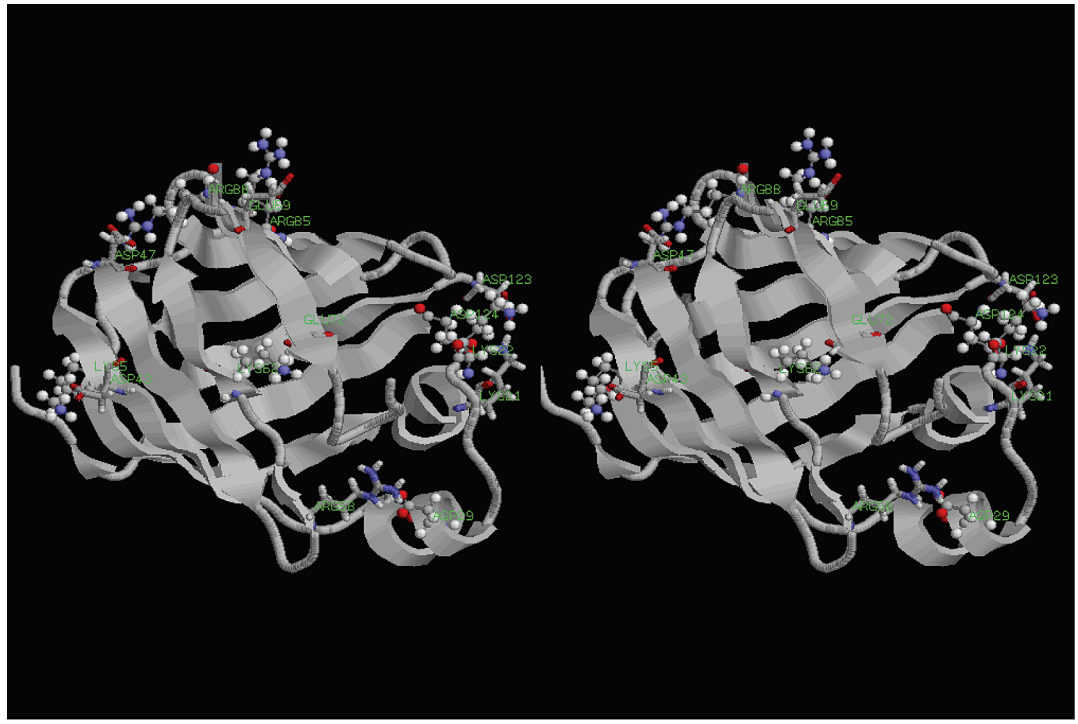

Figure 3. Stereoview of the mutated variant (designed to be thermophilic) of the protein *2rcq* with the key residues (Lys-21, Asp-43, Asp-47, Arg-58, Glu-69, Glu-72, and Asp-123) highlighted in the *stick* model and the mutated residues (Lys-5, Lys-22, Asp-29, Lys-62, Arg-85, Arg-88, and Asp-124) represented in the *ball and stick* model. All the mutated residues reside close to the surface of the protein and are delocalized.

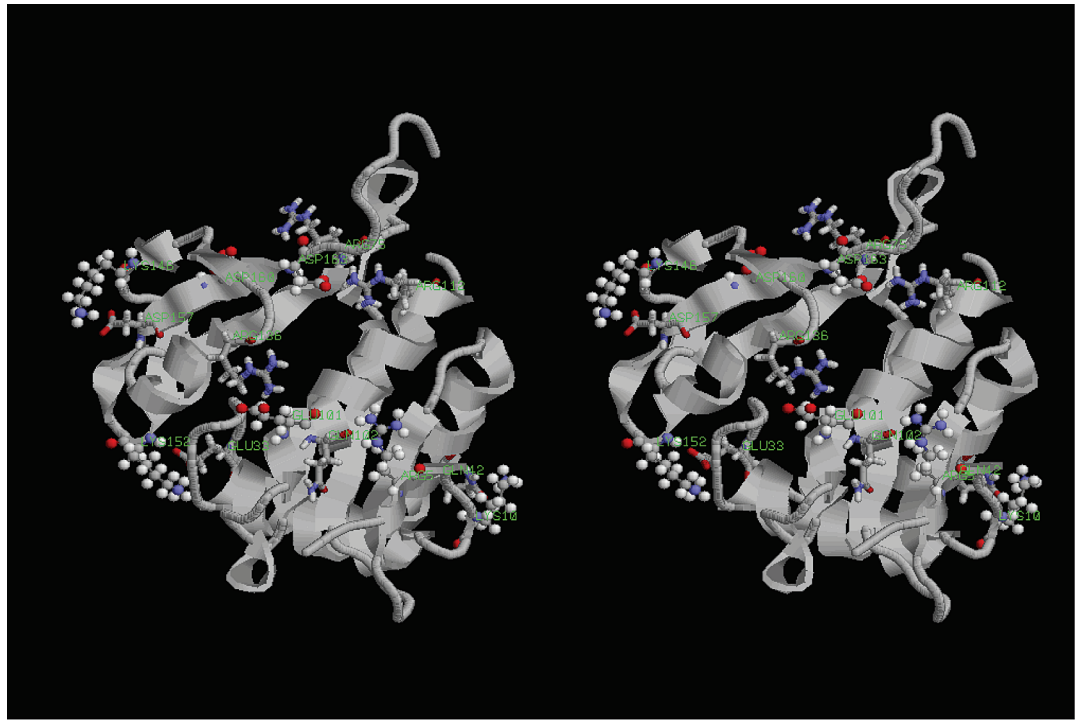

Figure 4. Stereoview of the mutated variant (designed to be thermophilic) of the protein *2bl1* with the key residues (Glu-33, Gln-42, Arg-75, Gln-102, Arg-112, Arg-136, and Asp-157) highlighted in the stick model and the mutated residues (Arg-5, Lys-10, Glu-101, Lys-146, Lys-152, As-p160, and Asp163) represented in the ball and stick model. The mutated residues are close to the surface of the protein and are delocalized over the protein.

Table 3. Comparison of the Self-Energies between the Known Mesophilic−Thermophilic Protein Pairs for the Validation of the Methods[a]

protein pair						
thermo	meso	sequence identity (%)	E_{self}^{meso} (kcal/mol)	E_{self}^{therm} (kcal/mol)	ΔE_{self} (kcal/mol)	ΔE_{self} *per residue* (kcal/mol)
1caa (53)	8rxn (52)	70.8	−93.6	−272.5	−178.9	−3.3
1yna (193)	1xnb (185)	49.5	−1585.6	−1872.0	−286.4	−1.1
1tmy (118)	3chy (128)	29.1	−1286.6	−1210.5	76.1	−0.2
2prd (174)	1ino (175)	47.4	−1224.0	−1816.7	−592.7	−3.4
1c9o (66)	1csp (67)	84.6	−437.1	− 561.0	−123.9	−2.0

[a] $\Delta E_{self} = E_{self}^{th} - E_{self}^{me}$ (kcal/mol). The number of residues in each protein is given within the brackets. The quantity ΔE_{self} *per residue* has been defined as ΔE_{self} *per residue* $= (E_{self}^{therm}/N_{therm}) - (E_{self}^{meso}/N_{meso})$, where N_{therm} and N_{meso} represent the number of residues in the thermophilic and the mesophilic variant, respectively. Thus, it is a useful quantity to compare the relative stability of proteins of different sizes.

we observed in Figures 1 and 2. All these features validate our strategy. As the sequence identities between the thermophilic and its mesophilic counterpart were not very high for some of the protein pairs, much noise was observed in the ΔE_i vs residue number plots (Figure 5b−d) arising from the nonconserved residues between them. It may be pointed out here that not all the nonconserved residues between the thermophilic−mesophilic protein pair are responsible for the observed thermostability, and most likely they have an evolutionary origin as these are naturally occurring thermophilic−mesophilic protein pairs.

Among the chosen protein pairs of the validation set, *1c9o* and *1csp* represent experimentally the most well studied protein pair[9] and are of almost equal length. So we have considered this protein pair in greater detail in validating our approach. There are 12 differences in the amino acid residues in the sequences between the thermophilic (*1c9o*) and mesophilic counterpart (*1csp*) as summarized in Table 4. We energy minimized the H-atom assigned respective crystal structures as mentioned before and computed their self-energies. Table 3 indicates that the computed ΔE_{self} was found to be −123.9 kcal/mol indicating significant increase in the stability of the thermophilic variant *1c9o*. This is quite consistent with experimental data that the melting temperature of *1c9o* is 23 °C higher than that of *1csp*.[9] We further observed features in the ΔE_i vs i plot (Figure 5e) for this protein pair similar to the features we observed in Figures 1 and 2. These results validate our strategy further.

Table 4 also summarizes the residue-wise best ΔE values for the nonmatching residue pairs of *1c9o* and *1csp* considering electrostatic and *van der Waals* interactions. It is seen that in several cases the most strongly interacting pairs are 'sequence-wise local', and thus these are not important for stabilizing or destabilizing the protein due to change in residues. Moreover, it is found that Arg-3 and Glu-46 are the residues that contribute maximally in stabilizing the protein through electrostatics. Evidences of strong contribution coming from Arg-3 are also available in the literature.[9] However, it is also believed that hydrophobic interactions associated with the residue Leu-66 are responsible partially for the stability of *1c9o*,[9] and it is not captured here as we only probe the interactions through electrostatics and *van der Waals* components. The +23.0 kcal/mol of the ΔE value resulting from the Glu-66→Leu-66 change only indicates that there is some residual steric conflict due to the van der Waals interaction.

DISCUSSIONS

In the present design strategy, we have paid the most attention to the electrostatic part of the inter-residue interac-

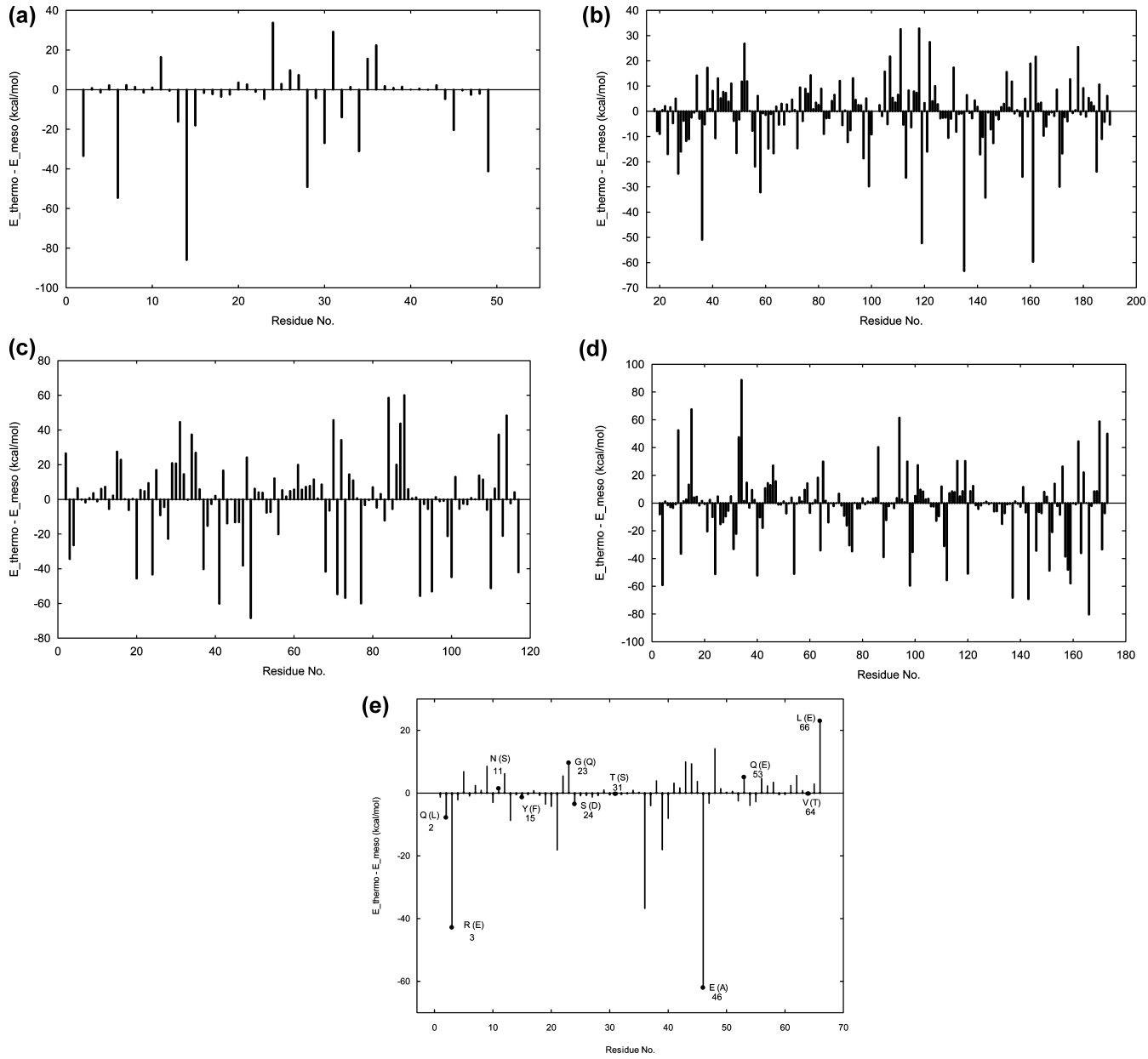

Figure 5. a: The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the thermophilic protein (*1caa*) and its existing mesophilic counterpart (*8rxn*) of the validation protein pair is shown. E_i^{Th} and E_i^{Me} represent the interaction energies of the ith residue of *1caa* and *8rxn*, respectively. **b:** The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the thermophilic protein (*1yna*) and its existing mesophilic counterpart (*1xnb*) of the validation protein pair is shown. E_i^{Th} and E_i^{Me} represent the interaction energies of the ith residue of *1yna* and *1xnb*, respectively. **c:** The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the thermophilic protein (*1tmy*) and its existing mesophilic counterpart (*1chy*) of the validation protein pair is shown. E_i^{Th} and E_i^{Me} represent the interaction energies of the ith residue of *1tmy* and *1chy*, respectively. **d:** The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the thermophilic protein (*2prd*) and its existing mesophilic counterpart (*1ino*) of the validation protein pair is shown. E_i^{Th} and E_i^{Me} represent the interaction energies of the ith residue of *2prd* and *1ino*, respectively. **e:** The pattern of the differences in interaction energies ($\Delta E_i = E_i^{Th} - E_i^{Me}$) of individual residues (ith one) of the mesophilic protein (*1csp*) and its existing thermophilic counterpart (*1c9o*) of the validation protein pair is shown. E_i^{Th} and E_i^{Me} represent the interaction energies of the ith residue of *1c9o* and *1csp*, respectively. The residues selected for mutations are indicated by a filled circle (•) symbol at the end of the respective bar along with the mutation indicated beside the bar with the original residue name within brackets. The residue numbers are also indicated.

tion. The effect of hydrophobic improvement is only partially taken care of by considering the van der Waals interactions, as close contact among hydrophobic side chains should enhance the van der Waals component of interactions.

We have considered only the enthalphic part of the free energy differences in designing the thermophilic protein as our objective is to increase the stability of the protein in a rather qualitative way. However, for a more accurate and quantitative description of stability, entropic contributions should also be used as entropic effects may produce additional favorable and unfavorable contributions. For considering the entropic contributions an ensemble of structures is necessary. In this regard molecular dynamics simulations of the mesophilic and thermophilic variants of the proteins in explicit solvent seem to be useful. Similarly, the solvation contribution can be incorporated through solvation free energy calculation using the Linear Interaction Energy (LIE) method[43] that may also be useful in identifying

Table 4. Differences in the Residues in the Mesophilic and Thermophilic Partners of the Validation Protein Pair[a]

no.	mesophile (CSP)	thermophile (C9O)	ΔE_i (kcal/mol)	best partner	Δres	D_{CA} (Å)
1	Leu-2	Gln-2	−7.9	Arg-3	1	3.8
2	Glu-3	Arg-3	−42.9	Glu-46	43	7.3
3	Ser-11	Asn-11	1.4	Thr-40	29	6.3
4	Phe-15	Tyr-15	−1.4	Asn-10	5	5.5
5	Gln-23	Gly-23	9.6	Ser-24	1	3.8
6	Asp-24	Ser-24	−3.6	Gly-23	1	3.8
7	Ser-31	Thr-31	−0.3	Phe-30	1	3.9
8	Ala-46	Glu-46	−62.1	Lys-5	41	4.4
9	Glu-53	Gln-53	5.0	Val-52	1	3.8
10	Thr-64	Val-64	−0.3	Ser-48	16	5.1
11	Glu-66	Leu-66	23.0	Lys-65	1	3.8
12	Ala-67					

[a] The differences in interaction energies of individual residues and their best interacting partner residues are also indicated. Δres represents the gap between the key residue and its best partner along the sequence in terms of a number of residues. D_{CA} represents the spatial distance (Å) between the C-alpha atoms of the key residue and its best (most strongly interacting) partner in the thermophilic variant.

the polar and charged amino acid residues that interact poorly with the rest of the protein. In our future work we will try to incorporate these aspects.

The present approach has several advantages over the other methods where sequence comparison plays the major role. The present approach is based on the actual 3D structure of the original protein. Furthermore, it takes care of realistic physical interactions between the individual residues at the atomic level and thus provides the opportunity to enhance the stability of the protein through the inter residue interactions in the 'spatially local' region. Finally, the consideration of the van der Waals interactions enables design avoiding steric conflicts among residues. On the other hand as it requires the 3D structure of the mesophilic protein for designing the thermophilic variant, in one sense it could be considered as the only disadvantage of the present approach.

In the case of designing a thermophilic version of some functional proteins, it is important not to mutate any of the residues involved in the active site as alteration in the active site should have adverse effects on its activity.

In order to enhance the stability further, the procedure can be repeated for a couple of times using the modified protein to increase the number of the favorable mutations.

The present procedure can even be applied to a region where no polar or charged residue is found. In such a case, some preselected residue in the region of interest should be replaced by suitable polar or charged residue following our methods. Subsequently, some residue in its vicinity should be identified and replaced by a side chain of complementary polar nature as described in this work.

Thus, depending on the amino acid sequence and composition, some proteins may be turned into a thermophilic one even in a single cycle of the entire procedure described here. In principle, several consecutive cycles of the present procedure should be able to turn any protein into a thermophilic one.

CONCLUSIONS

The present method of designing a thermophilic protein starting with the 3D structure of the respective mesophilic protein has been demonstrated to be quite effective in the cases considered here. The method appears to be very simple and fast but robust and applicable to any kind of protein. Consideration of the 3D structure has made the approach more straightforward and accurate in designing thermophilic proteins.

ACKNOWLEDGMENT

The authors would like to thank Ms. Nandita Sinha for occasional discussions and critically reading the manuscript.

Supporting Information Available: List of the PDB IDs for the protein structures we have used in generating the conformers libraries. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Jaenicke, R. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2926–2964.
(2) Szilagyi, A.; Zavodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comparative survey. *Structure* **2000**, *8*, 493–504.
(3) Fitter, J.; Heberle, J. Structural equilibrium fluctuations in mesophilic and thermophilic -amylase. *Biophys. J.* **2000**, *79*, 1629–1636.
(4) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1–43.
(5) Vogt, G.; Woell, S.; Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **1997**, *269*, 631–643.
(6) Spassov, V. Z.; Karshikoff, A. D.; Ladenstein, R. The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci.* **1995**, *4*, 1516–1527.
(7) Kumar, S.; Nussinov, R. How do thermophilic proteins deal with heat. *Cell. Mol. Life Sci.* **2001**, *58*, 1216–1233.
(8) Missimer, J. H.; Steinmetz, M. O.; Baron, R.; Winkler, F. K.; Kammerer, R. A.; Daura, X.; Van Gunsteren, W. F. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci.* **2007**, *16*, 1349–1359.
(9) Perl, D.; Mueller, U.; Heinemann, U.; Schmid, F. X. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* **2000**, *7*, 380–3.
(10) Park, S.; Xu, Y.; Stowell, X. F.; Gai, F.; Saven, J. G.; Boder, E. T. Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng. Des. Sel.* **2006**, *19*, 211–7.
(11) Razvi, A.; Scholtz, J. M. Lessons in stability from thermophilic proteins. *Protein Sci.* **2006**, *15*, 1569–78.
(12) Fitter, J.; Heberle, J. Structural equilibrium fluctuations in mesophilic and thermophilic -amylase. *Biophys. J.* **2000**, *79*, 1629–1636.
(13) Thomas, A. S.; Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J. Am. Chem. Soc.* **2004**, *126*, 2208–2214.
(14) Gribenko, A. V.; Makhatadze, G. I. Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J. Mol. Biol.* **2007**, *366*, 842–56.
(15) Grimsley, G. R.; Shaw, K. L.; Fee, L. R.; Alston, R. W.; Huyghues-Despointes, B. M.; Thurlkill, R. L.; et al. Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **1999**, *8*, 1843–1849.
(16) Perl, D.; Schmid, F. X. Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **2001**, *313*, 343–57.
(17) Dominy, B. N.; Minoux, H.; Brooks, C. L., 3. An electrostatic basis for the stability of thermophilic proteins. *Proteins* **2004**, *57*, 128–41.
(18) Karshikoff, A.; Ladenstein, R. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **1998**, *11*, 867–72.
(19) Vogt, G.; Argos, P. Protein thermal stability: hydrogen bonds or internal packing. *Fold. Des.* **1997**, *2*, S40−6.
(20) Xiao, L.; Honig, B. Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **1999**, *289*, 1435–1444.
(21) Karshikoff, A.; Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends. Biochem. Sci.* **2001**, *26*, 550–556.
(22) Elcock, A. H. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **1998**, *284*, 489–502.

(23) Danciulescu, C.; Ladenstein, R.; Nilsson, L. Dynamic Arrangement of Ion Pairs and Individual Contributions to the Thermal Stability of the Cofactor-Binding Domain of Glutamate Dehydrogenase from *Thermotoga maritima*. *Biochemistry* **2007**, *46*, 8537–8549.
(24) Tanner, J. J.; Hecht, R. M.; Krause, K. L. Determinants of Enzyme Thermostability Observed in the Molecular Structure of *Thermus aquaticus* D-Glyceraldehyde-3-phosphate Dehydrogenase at 2.5 Å Resolution. *Biochemistry* **1996**, *35*, 2597–2609.
(25) Montanucci, L.; Fariselli, P.; Martelli, P. L.; Casadio, R. Predicting protein thermostability changes from sequence upon multiple mutations. *Bioinformatics* **2008**, *24*, i190−i195.
(26) Lehmann, M.; Pasamontes, L.; Lassen, S. F.; Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* **2000**, *1543*, 408–415.
(27) Lehmann, M.; Loch, C.; Middendorf, A.; Studer, D.; Lassen, S. F.; Pasamontes, L.; van Loon, A. P.; Wyss, M. The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **2002**, *15*, 403–11.
(28) Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **2005**, *308*, 857–860.
(29) Zollars, E. S.; Marshall, S. A.; Mayo, S. L. Simple electrostatic model improves designed protein sequences. *Protein Sci.* **2006**, *15*, 2014–2018.
(30) Dantas, G.; Kuhlman, B.; Callender, D.; Wong, M.; Baker, D. A Large Scale Test of Computational Protein Design: Folding and Stability of Nine Completely Redesigned Globular Proteins. *J. Mol. Biol.* **2003**, *332*, 449–460.
(31) Loladze, V. V.; Ibarra-olero, B.; Sanchez-Ruiz, J. M.; Makhatadze, G. I. Engineering a thermostable protein via optimization of charge-charge interactions on protein surface. *Biochemistry* **1999**, *38*, 16419–16423.
(32) Spector, S.; et al. Rational modification of protein stability by mutation of charged surface residues. *Biochemistry* **2000**, *39*, 872–879.
(33) Strickler, S. S.; et al. Protein stability and surface electrostatics: a charged relationship. *Biochemistry* **2006**, *45*, 2761–2766.
(34) Strub, C.; Alies, C.; Lougarre, A.; Ladurantie, C.; Czaplicki, J.; Fournier, D. Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochem.* **2004**, *5*, 9, doi:10.1186/1471-2091-5-9.
(35) Loladze, V. V.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Makhatadze, G. I. Engineering a thermostable protein via optimization of charge−charge interactions on the protein surface. *Biochemistry* **1999**, *38*, 16419–16423.
(36) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. *J. Mol. Biol.* **2004**, *336*, 929–942.
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minmimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
(38) MacKerell, A. D., Jr.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(39) Sen, S.; Nilsson, L. Structure, interactions, dynamics and solvent effects of the DNA-EcoRI complex in aqueous solutions from Molecular Dynamics simulation. *Biophys. J.* **1999**, *77*, 1782–1800.
(40) Sen, S. Exploring Structure and Energetics of a Helix forming oligomer by MM and MD simulation Methods: Dynamics of water in hydrophobic nanotube. *J. Phys. Chem. B* **2002**, *106*, 11343–11350.
(41) Roy, S.; Sen, S. Homology Modeling based solution structure of Hoxc8-DNA complex: role of context bases outside TAAT stretch. *J. Biomol. Struct. Dynam.* **2005**, *22*, 707–718.
(42) Roy, S.; Sen, S. Exploring the potential of complex formation between a mutant DNA and the wild type protein counter part: A MM and MD simulation approach. *J. Mol. Graphics Modell.* **2006**, *25*, 158–168.
(43) Bren, U.; Martinek, V.; Florian, J. Free Energy Simulations of Uncatalyzed DNA Replication Fidelity: Structure and Stability of T·G and dTTP·G Terminal DNA Mismatches Flanked by a Single Dangling Nucleotide. *J. Phys. Chem. B* **2006**, *110*, 10557–10566.

CI900183M