Journal of Molecular Graphics and Modelling 42 (2013) 92–103

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

An *in silico* method for designing thermostable variant of a dimeric mesophilic protein based on its 3D structure

Sohini Basu, Srikanta Sen*

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

ARTICLE INFO

Article history:
Received 1 December 2012
Received in revised form 25 February 2013
Accepted 27 February 2013
Available online 21 March 2013

Keywords:
Protein design
Protein dimer
Mutation
Conformer library
Self-energy
Stability

ABSTRACT

Designing proteins with enhanced thermostability has been a major interest of protein engineering because of its potential industrial applications. Here, we have presented a computational method for designing dimeric thermostable protein based on rational mutations on a mesophilic protein. Experimental and structural data indicate that the surface stability of a protein is a major factor controlling denaturation of a protein and ion-pairs are most efficient in enhancing the stability of the surfaces of the monomers and the interface between them. Our mutation based strategy is to first identify several polar or charged residues on the protein surface, interacting weakly with the rest of the protein and then replacing the side-chains of suitable neighboring residues to increase the interaction between these two residues. In stabilizing the interface, mutation is done in the interface for forming an ion pairs between the monomers. Application of this design strategy to a homo-dimeric protein and a hetero-dimeric protein as examples has produced excellent results. In both the cases the designed mutated proteins including the individual monomers and the interfaces were found to be considerably more stable than the respective mesophilic proteins as judged by self-energies and residue-wise interaction patterns. This method is easily applicable to any multi-meric proteins.

1. Introduction

Organisms whose optimal growth temperatures are in the range from 20 to 50 °C are known as mesophilic organisms. Mesophilic organisms are most common in nature. Over the last decade, new organisms have been found in hot places such as hot springs and hot vents at the bottom of the sea. Such organisms grow and function at high temperatures and are classified as thermophiles [1,2]. If the optimal growth temperature (T_{opt}) of the thermophilic organism is in the range 50 °C < T_{opt} < 80 °C, the organism is called a moderate thermophile while, organisms with T_{opt} > 80 °C, are called hyperthermophiles. Proteins isolated from thermophilic and hyperthermophilic organisms remain structurally stable and functionally active at much higher temperatures at which their mesophilic counterparts are structurally unstable and hence inactive [3,4]. Thermophilic and hyperthermophilic proteins and their mesophilic counterparts generally possess high degree of sequence similarity and similar three-dimensional structures [5,6].

Designing proteins with enhanced thermostability is one of the major focuses of modern protein engineering. Thermophilic proteins are of great interest not only for their extraordinary stability but also because of their potential industrial uses [2,7–9]. Thus, it is important to understand the physical basis of the enhanced stabilities of thermophilic proteins at elevated temperatures and, also to use that acquired knowledge in designing tailor-made thermostable proteins for industrial applications. There are a few approaches for making a given protein a thermophilic one [10–14]. However, these methods depend mostly on the sequence comparison, 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. Thus, designing methods based on the 3D structure of the protein appears to be more straight forward and reliable. In the present work we have proposed a structure-based effective and robust computational approach for converting a mesophilic multimeric protein into a thermophilic one. This method is an extension of our earlier design approach for turning a single chain single globule mesophilic protein into a thermophilic variant [15,16].

The physical origin of the enhanced stability of thermophilic proteins has been investigated extensively [16–25]. Our current understanding indicates that there are a number of factors that individually or in combination can cause the extra stability. Such factors include optimized electrostatic interactions such as increased number of salt-bridges, improved packing, networks of hydrogen bonds, increased hydrophobic interactions, and decreased number and volume of internal cavities, *etc.* [17–33]. Till now, no general rules have been identified to determine thermostability, and different thermophilic proteins appear to achieve their increased

* Corresponding author. Tel.: +91 33 23572744/45; fax: +91 33 23574113.
E-mail addresses: srikantasen@ymail.com, srikanta@chembiotek.com (S. Sen).

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2013.02.011

thermostability due to different combinations of the above mentioned factors. In the natural systems the enhancement of stability due to multiple factors is more likely. As a consequence, in designing thermophilic variants of mesophilic proteins one can prefer a single factor or a few factors of choice. However, it is now well accepted that electrostatic interactions like ion-pairs are the most common factors observed in hyper-thermophilic proteins. Comparative studies have revealed that many thermophilic proteins have increased number of salt bridges [1,17,22,23,25] compared to the homologous mesophilic proteins. Moreover, a large number of salt bridges are generally found to be present on the surface of thermophilic proteins, and these salt bridges may participate not only in intra-protein interactions but also in inter-domain and inter-subunit interfaces of proteins for the stabilization of the whole protein. In a recent paper, it has been reported that a protein CutA1 isolated from *Pyrococcus Horikoshii*, has a very high melting temperature (150 °C) and its crystal structure indicates that the surface of this protein is almost covered by ion-pairs [25]. Based on these facts, we have developed a computational method for generating thermophilic proteins by introducing mutations that improve the electrostatic energy of the 3D structure of the respective mesophilic counterpart. We have published earlier the main concept and its validation along with the demonstration of designing thermostable variants of single chain single globule mesophilic proteins [15,16]. The purpose of the present work is to demonstrate that the same approach can be extended to convert a dimeric mesophilic protein into a dimeric thermophilic protein. It is important to emphasize that in converting a multimeric mesophilic protein into a multimeric thermostable protein, it is not enough to simply enhance the stabilities of the individual monomers only, but also to stabilize further the interfaces between the monomers such that the quaternary complex is stabilized as a whole. For the sake of simplicity, here we have demonstrated two examples of dimeric proteins of which, one is a homo-dimer and the other is a hetero-dimer but the approach can be applied to any multimeric protein as well. In both the examples it has been demonstrated explicitly that the designed mutations have improved the stability of the individual monomers as well as the stability of the interface. As experimental verification for the enhanced stability of the designed protein is not amenable to us, we have demonstrated the enhanced stability in terms of computational properties. We have chosen the modulation of electrostatic interactions to achieve enhanced stability because, so far, it is known to be the most common stabilizing factor in many cases. Consideration of other factors in designing thermophilic proteins may be addressed in the future. The present method is found to be a versatile and straight forward design approach in converting a multimeric mesophilic protein into a multimeric thermostable one. Validation of the general approach of our design method based on rational mutations in converting a single chain single globule mesophilic protein into a thermophilic protein has already been described in our earlier work where, it was demonstrated by considering several naturally occurring homologous thermophilic and mesophilic protein pairs that such natural protein pairs satisfy all the criteria we have used in converting a mesophilic protein into a thermophilic one [15,16]. In the present work the design method has been extended to cover dimeric proteins through additional stabilization of the interface.

2. Materials and methods

2.1. Rational for stabilizing the interface at high temperature

It is known that hydrophobic interaction plays an important role in forming dimeric proteins. However, as hydrophobic interaction is weakened with increase in temperatures, it is unlikely that hydrophobic interactions provide enhanced stability of the interface for a thermophilic protein. On the other hand, there are evidences that the interfaces for dimeric thermostable proteins are stabilized by ion pairs formed between the monomers [34–36]. Based on this fact, we have developed a method for identifying mutations on the two monomers at the interface allowing the formation of inter-monomer ion pairs and thus, stabilizing the dimer.

2.2. Choice and preparation of the dimeric mesophilic proteins of interest

The RCSB database was searched to find the crystal structures of dimeric proteins of moderate sizes and a homo-dimeric protein (*Fibroblast growth factor receptor 2*, PDB ID: 3EUU) and another hetero-dimeric protein (*Sugar binding protein Lectin 1*, PDB ID: 1GGP) were selected as the 'test systems'. We prepared each protein as follows. The homo-dimeric crystal structure 3EUU contains 100 residues for each monomer. First of all H-atoms were assigned to the crystal structure of the selected protein using CHARMM. Then the 3D structure was refined by performing energy minimization by 5000 steepest descent steps in vacuum. This removed bad steric contacts from the structure if any. The hetero-dimeric protein 1GGP contains 234 residues in chain A and 254 residues in chain B and was prepared in the same way. These two protein structures were then used separately to design their thermostable variants.

2.3. Generation of side-chain conformers libraries

We generated the conformer libraries for the 19 different kinds of amino acid residues directly from 49 PDB files randomly selected from the RCSB database and using our in-house tool [15,16]. We have not considered Gly in this library as it contains only one H-atom as its side-chain. We visually inspected the generated conformers and found it to represent the conformational space available to the side-chain reasonably well. The conformer models also cover their diverse orientations satisfactorily. The advantage of this method is that the conformers of the side-chains in all the cases are experimentally found ones.

2.4. Designing of more stable protein dimer

The entire designing process has two major steps (1) designing of more stable individual monomers and (2) designing of more stable monomer–monomer interface.

2.4.1. Designing of more stable individual monomers

The stability of an individual monomer is enhanced by introducing rational based mutations in it. The rational and the detailed method have been discussed earlier [15,16]. Here we are giving the essential summary of the approach. Based on our current understanding of the physical origin of the enhanced stability of thermophilic proteins we have selected to modulate the electrostatic component of the self-energy of the protein by mutating a set of rationally selected residues on or close to the surface of the protein. We have chosen mutations on or close to surface as such mutations are reported to enhance thermostability [25–33]. Moreover, it is expected that stabilization of the structural fluctuation of the surface will reduce the possibility of the onset of denaturation. Further more, mutations at the surface are less likely to alter the folded structure. The details of the original approach of converting a single chain mesophilic protein into a designed thermostable one are available in our earlier publications [15,16]. Fig. 1 shows the summary of the schematic diagram of the work flow. The steps of the present design approach are summarized below.

Step-1: First, the energy-minimized 3D structure of the original protein was considered and the interaction energies of the

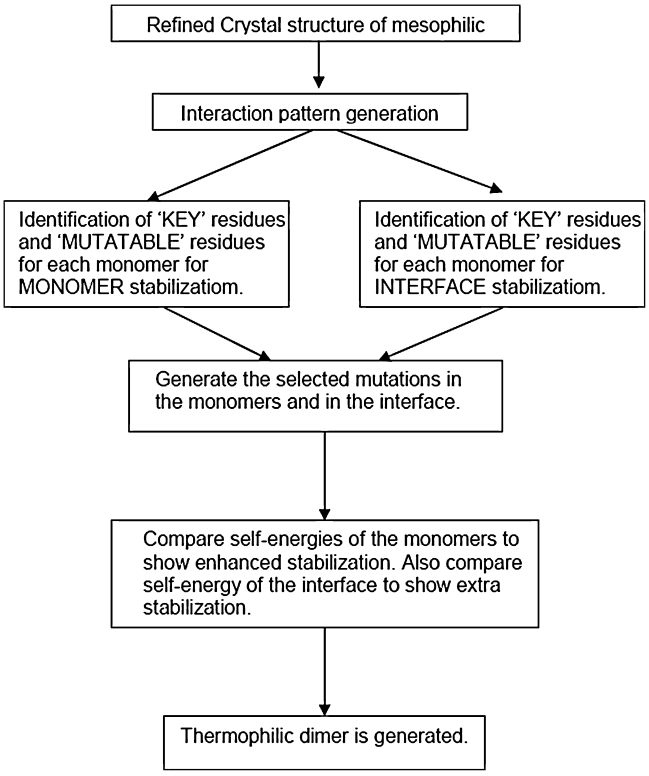

Fig. 1. The schematic work flow of the computational design method described in this manuscript.

side-chain of its individual residues with the rest of the protein were generated by considering only the electrostatic and van der Waals interactions.

The electrostatic interaction energy was computed by standard Coulomb's law

$$E_{\text{elec}} = 331.5 \times \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{q_i q_j}{\varepsilon . d_{ij}}$$

where q_i represents the partial atomic charge of the ith atom of the respective side-chain of the residue of interest, and q_j represents the same for the atoms of the rest of the protein. We used a spherical distance cut-off value of 12.0 Å for computing the non-bonded interactions and a distance dependent dielectric constant ($\varepsilon = \varepsilon_0 d_{ij}$) as in CHARMM. It may be mentioned here that even though we have used a distance dependent dielectric constant in computing the electrostatic component of the inter-residue interaction, any other value of dielectric constant would scale the electrostatic energy value only but, the relative values of the electrostatic energies of the different residues would be maintained.

The Van der Waals interaction energy was computed using the Lennard–Jones formula,

$$E_{\text{vdw}} = \sum_{i=1}^{n} \sum_{\substack{i=1 \\ j \neq i}}^{n} e_{ij} \left[\left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{d_{ij}} \right)^{6} \right]$$

where $\sigma_{ij} = (1/2)(\sigma_i + \sigma_j)$, σ_i and σ_j representing the Lennard–Jones diameters of the ith and jth atoms respectively, d_{ij} corresponds to the inter-atomic distance between the ith and jth atoms, and $e_{ij} = \sqrt{e_i e_j}$, with e_i and e_j being the Lennard–Jones potential well-depth of the ith and jth atoms respectively. We used the partial atomic charges and the Van der Waals parameters following CHARMM topologies and parameters.

The residue-wise interaction pattern was then examined in order to identify the polar or charged residues whose side-chains interacting weakly with the rest of the protein. Such residues were then considered as the 'key residues'.

Step-2: We rejected all the key residues that were not close to the surface. Subsequently, for each remaining key residue, we identified some other nearby residue that was interacting weakly with the selected key residue as well as with the rest of the protein. We call such a residue a 'mutatable residue' and its side-chains were chosen for replacement by suitable side-chains from the 'side-chain conformer library' in order to improve its interaction with the respective key residue and hence the stability of the protein. In selecting the 'key-residue' and its partner 'mutatable 'residue' care was taken such that they were not close along the sequence because, in that case the increase in their mutual interaction will not be effective in improving the stability. We considered a minimum cut off value of 10 residues for the separation between them along the sequence.

Step-3: We have developed in-house computer 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. In the selection process, for each mutatable residue, all the individual side-chain conformers with different spatial orientations were substituted, and its interaction energy with the rest of the protein was computed as explained before and the side-chain conformer giving the strongest energy of interaction with the rest of the protein was identified as a potential substitution. This step was repeated for all other 18 (excluding Gly and self type) residues. We select the best replacement considering the most favorable interaction energy with the rest of the protein and replace the original side-chain by it. These steps were repeated to replace all the selected mutatable side-chains one by one in a cumulative fashion. At the end of the process we obtained the atomic coordinates of the designed protein with the substituted side-chains at the selected residues.

Once the mutations in both the monomers were done we moved to the steps of stabilizing the interface.

2.4.2. Designing of more stable monomer–monomer interface

The stability of the monomer–monomer interface was enhanced by suitable mutations in the interface such that it improved its interaction with other monomer at the interface. Thus, mutation on either monomer or both can be useful. In this phase, our starting point was the dimeric protein with mutated monomers obtained in the previous section. The next steps of the process are summarized below.

Step-4: The residue-wise interaction pattern between the two monomers was first computed. The interaction pattern indicated the residues that were in the interface. In one monomer, the polar or charged residues that were interacting weakly with the other monomer were identified as the 'key residues'.

Step-5: The 'mutatable residue' corresponding to each key residue in the monomer was identified on the other monomer from the interaction pattern. For enhancing the stability of the interface, the chosen mutatable residue and the respective key residue must not be on the same monomer. Otherwise, stabilization of the interface will not occur.

Step-6: As explained before in the case of individual monomers, the mutatable residues were substituted by the most suitable residues as obtained from the conformer library monitoring its interaction with the rest of the protein including both the monomers as described before. In order to ensure the stabilization of the interface, the interaction energy of the substituted side-chain with the other monomer was checked. Its interaction with the same

monomer was also computed to check that the substitution did not create any bad contact within the same monomer. The process was repeated for all the mutatable residues in a cumulative way where all the previous mutations were included. It is desirable to make such mutations on both the monomers to keep the degree of change small in a single monomer.

Step-7: Once, all the mutatable residues were replaced, the resulting protein structure was refined by energy minimization using CHARMM considering the full force field [37,38].

The stability of a protein is defined as the difference in free energies between its folded and the unfolded forms. In the present case, it may be pointed out that the mutated protein is different in its sequence and amino acid compositions from the original protein. However, as the number of such mutations is only a few compared to the whole protein, the difference in the free energies between the original and mutated proteins in their unfolded states may be neglected for simplicity. In that case, the difference in self-energies between the original protein and the mutated proteins in their folded states may be an approximate measure of the difference in stabilities of the two proteins. The self-energy (E_{self}) of a dimer protein in a given 3D structure represents the total energy content of the structure including all the bonded and non-bonded interaction energy components of the individual monomers as well as the interactions between the two monomers. Thus, the difference ($\Delta E_{self} = E_{self}^{Mutated} - E_{self}^{Meso}$) in the self-energies of the mutated ($E_{self}^{Mutated}$) and mesophilic versions (E_{self}^{Meso}) 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 mutated protein is more stable than the corresponding original one. A more detailed discussion on the dependence of 'Stability' and 'Self-energy' of proteins is available in the 'Supplementary Information'. However, for a more accurate estimation of the change in stability one should compute the free energy differences between the original and the mutated proteins taking into account the change in enthalpy and also the change in entropy due to the mutations. We will address this in our future work.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2004.08.011.

The above seven steps may be repeated for a few cycles to make the stability enhancement more pronounced if required.

The validation of the main concept of our approach was demonstrated earlier by considering several natural homologous thermoplilic and mesophilic protein pairs that such natural protein pairs satisfy all the criteria that we used in converting a single chain mesophilic protein into a thermophilic one [15,16]. The present work extends our earlier method to cover dimeric proteins as the simplest of multimeric proteins.

3. Results

In this work we have demonstrated the effectiveness of our method of designing thermostable proteins that are dimeric. We considered two examples where, in one example, (A) a homo-dimeric mesophilic protein was converted to a homo-dimeric thermostable protein and in the other (B) a hetero-dimeric mesophilic protein was converted to a hetero-dimeric thermostable protein following the approach presented here.

3.1. Converting a homo-dimeric mesophilic protein (3EUU) into a dimeric thermostable protein

In the case of a homo-dimeric mesophilic protein, one needs to take extra care in selecting the mutation in both the monomers, such that finally, the two chains remain identical as far as sequence is concerned. The monomer structures even in a homo-dimer crystal structure were found to be slightly different and for this, we performed the mutation process separately for the two individual monomers.

First of all we assigned H-atoms to the protein in the crystal structure 3EUU. It was subsequently energy minimized by CHARMM using 5000 steepest descent (SD) steps in vacuum. The residue-wise pattern of the interactions of the side-chains for the first monomer with the rest of the first monomer was computed using our computer codes. Fig. 2 represents the interaction pattern. It shows that the side-chains of the residues Lys-176, Asn-194, Asp-247 and Asn-228 are interacting rather weakly with the rest of the first monomer. Examination shows that all these residues are on or

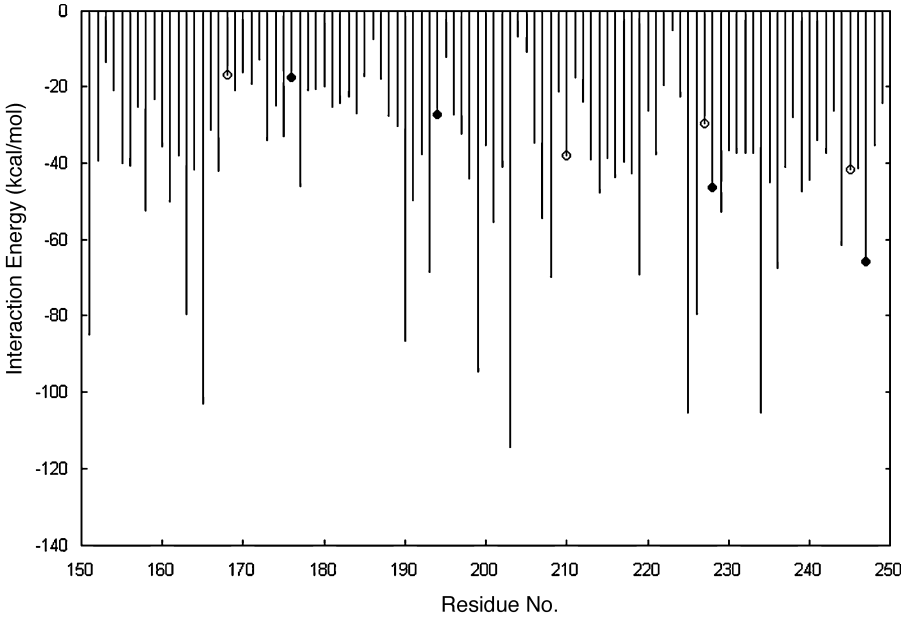

Fig. 2. Plot of the pattern of the residue-wise interaction energies of the individual residues of chain-A (monomer-1) of the energy minimized homo-dimeric mesophilic protein (3EUU) with the rest of the monomer-1. The filled circles represent the key residues and the selected mutatable residues are marked by open circles.

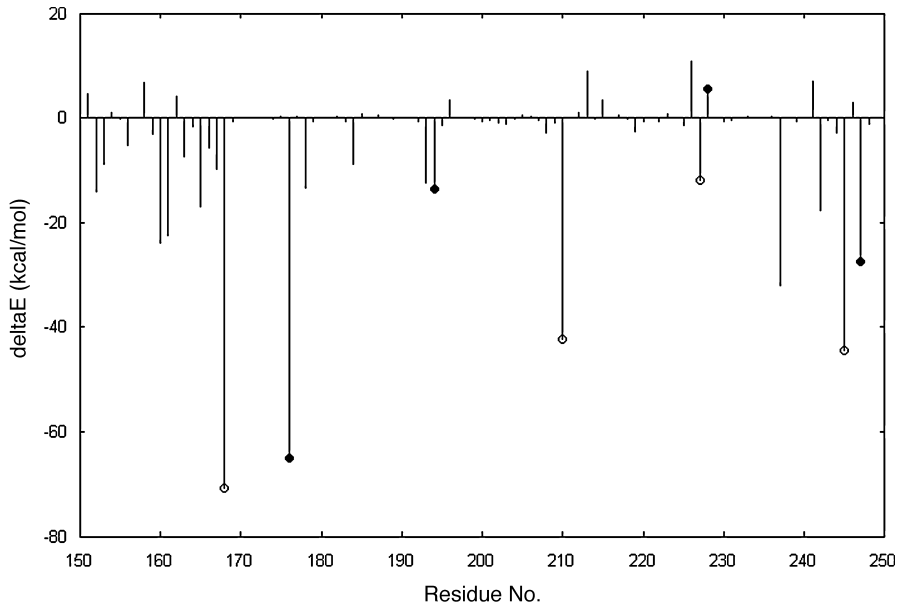

Fig. 3. The pattern of the residue-wise differences $\Delta E_i = E_i^{\text{Mutated}} - E_i^{\text{Meso}}$ in the interaction energies of the individual residues of chain-A with the rest of the chain-A between the original protein and the designed mutated protein in the case of the crystal structure 3EUU. The key residues and the mutated residues are highlighted by filled circles and open circles mark the mutatable residues.

close to the surface of the first monomer and are polar or charged. We considered these residues as the 'key residues' as mentioned in the Section 2. Further, it was found that the residues Arg-210, Gly-227, Ala-168 and His-245 are residues close to the individual key residues respectively. Thus, we considered primarily these residues for mutation. In choosing the mutatatble residues we also kept in mind that the 'key residue' and the 'mutatable residue' were not very close (separation < 10 residues) along the sequence as such cases will not affect the global stability of the monomer. The conformers of all the residues from the conformer libraries were tried for each selected mutatable residue and finally the most suitable mutations were found as Arg-210 → Glu-210, Gly-227 → Asp-227, Ala-168 → Lys-168 and His-245 → Glu-245. In a similar way the other monomer was also mutated. We mutated the monomers separately as the monomer structures even in a homo-dimer crystal structure were found to be slightly different.

Fig. 3 represents $\Delta E_i (= E_i^{\text{Mutated}} - E_i^{\text{Meso}})$ versus residue number i of the first monomer. It is seen that there are a number of large favorable changes in the residue-wise interaction in the mutated protein. It is found that many of the key residues have much favorable interaction energies in the mutated protein as originally designed. However, a few other residues are now showing improved interaction energies due to mutations made in their neighborhood. Similar data for the other monomer were also computed separately but are not shown as it is a homo-dimer.

The next step was to stabilize the interface by introducing rational mutations. Fig. 4 represents the residue-wise interaction pattern between the two monomers. Clearly, the residues

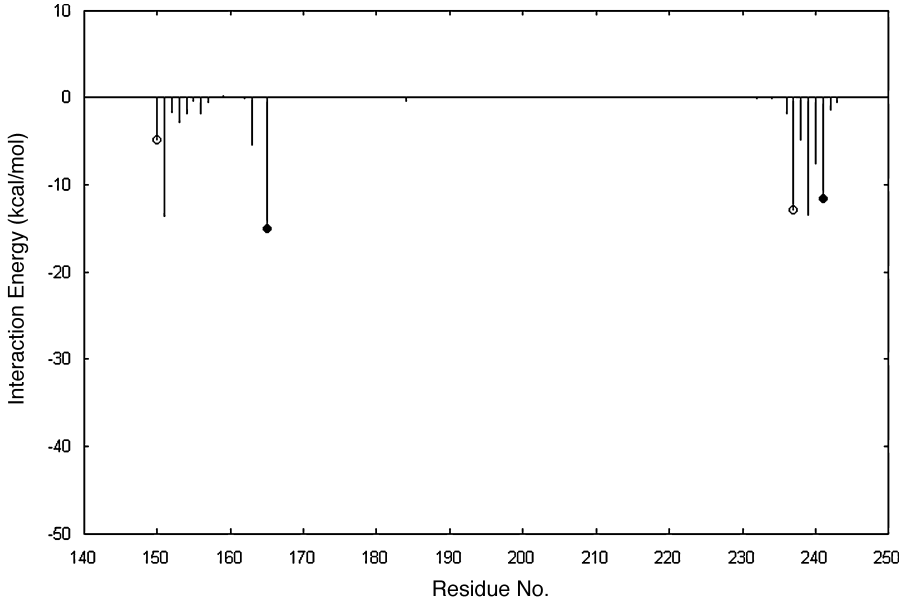

Fig. 4. The pattern of the interaction energies of the individual residues of monomer-1 (chain-A) with the other monomer (chain-B) in the homo-dimeric protein (3EUU) against the residue no. of chain-A. The residues with non-zero interaction energy are situated in the interface (AB). The residues chosen key residues are marked by filled circles and open circles represents the mutatable residues.

Table 1A
Summary of the results on the key residues, mutations, separation of the key residue and the mutatable residue along sequence, gain in energy due to mutation, *etc*., in the case of the protein with PDB ID: 3EUU.

Chain	No.	Key residue	Mutatable residue	Δres	Mutation made	ΔE_i[a] (kcal/mol)
A	1	Lys-176	Arg-210	34	Glu-210	−42.17
	2	Asn-194	Gly-227	33	Asp-227	−12.0
	3	Asn-228	His-245	17	Glu-245	−44.49
	4	Asp-247	Ala-168	79	Lys-168	−70.76
B	1	Lys-176	Arg-210	34	Glu-210	−39.5
	2	Asn-194	Gly-227	33	Asp-227	−12.11
	3	Asn-228	His-245	17	Glu-245	−13.33
	4	Asp-247	Ala-168	79	Lys-168	−40.10
Interface AB	1	Arg-165 (chain A)	Asn-150 (chain B)	NA	Glu-150 (chain B)	2.49
	2	Asn-241 (chain A)	Tyr-237 (chain B)		Glu-237 (chain B)	−27.57
Interface BA	1	Arg-165 (chain B)	Asn-150 (chain A)	NA	Glu-150 (chain A)	1.5
	2	Asn-241 (chain B)	Tyr-237 (chain A)		Glu-237 (chain A)	−27.43

[a] ΔE_i $(= E_i^{\text{ther}} - E_i^{\text{meso}})$, where E_i^{ther} is the interaction energy of the *i*th residue with the rest of the protein in the designed thermostable variant and E_i^{meso} is its mesophilic counterpart.

of monomer chain-A which correspond to non-zero interaction with the other monomer, were in the interface. It is seen that the residues Arg-165 and Asn-241 are interacting weakly with the other monomer and thus are suitable as the key residues for the interface. Following the same approach as mentioned before we found the residues Asn-150 and Tyr-237 in chain-B as the suitable mutatable residues. Then we replaced the side-chains of these mutatable residues following the same approach as mentioned before. This time we monitored the interactions of the substituted side-chain separately with the other monomer and also the same monomer where the mutation was done. This is because the interaction of the substituted side-chain with the other monomer provides the cross monomer interaction and it directly contributes to the stability of the interface while the interaction of the substituted side-chain with the same monomer where the mutation was done is important to ensure that the substitution is not involved in bad steric conflict with the same monomer where the substitution was made. These steps were repeated to replace both the selected mutatable side-chains one by one in a cumulative way. The best mutations were found as Asn-150 → Glu-150 and Tyr-237 → Glu-237. The details of the key residues and mutation data are summarized in Table 1A.

Fig. 5 represents the $\Delta E_i (= E_i^{\text{Mutated}} - E_i^{\text{Meso}})$ versus residue number *i* of the first monomer. It is again seen that there are a number of large favorable changes in the residue-wise interaction in the mutated protein. It is found that apart from the key residues, a few other residues also indicate growth of significant favorable interaction energies at the interface in the mutated protein as found in the cases of the individual monomers. These additional improved interactions were generated due to interactions between the mutated residues and the residues in their immediate neighborhood.

At the end of the mutation process we obtained a mutated protein which was expected to be more stable compared to the original protein. In order to check that, the generated structure of the new mutated protein dimmer was refined by energy minimization in vacuum following the same protocol as mentioned before and the self-energies and the residue-wise interaction patterns of the individual monomers as well as between the two monomers of the mutated protein were computed and compared to those of the original energy minimized mesophilic protein. Table 1B summarizes

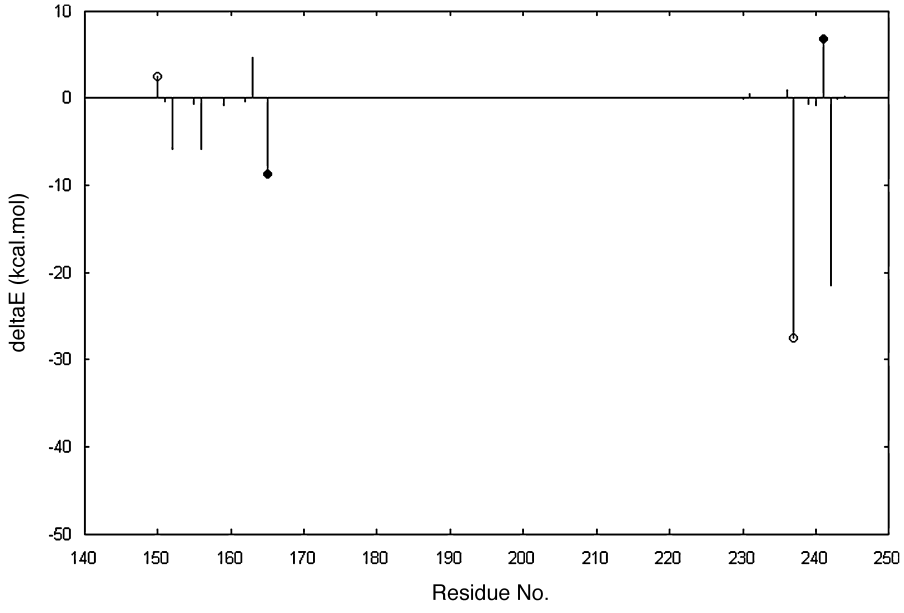

Fig. 5. The pattern of the residue-wise differences $\Delta E_i = E_i^{\text{Mutated}} - E_i^{\text{Meso}}$ in the interaction energies between the individual residues in the original protein and the designed mutated protein at the interface (AB) of the homo-dimeric protein 3EUUunder consideration. The key residues and the mutated residues are highlighted by filled circles and open circles mark the mutatable residues.

Table 1B
Comparison of the self-energies between the original mesophilic protein (PDB ID: 3EUU) and its designed mutated thermostable variant.

	Self energy (E^{ms}) of the mesophilic protein (kcal/mol)	Self energy (E^{mu}) of the designed protein (kcal/mol)	$\Delta E = (E^{mu} - E^{ms})$ (kcal/mol)	ΔE/res (kcal/mol/res)
Chain A	−996.82	−1077.48	−80.65	−0.81
Chain B	−992.06	−1067.42	−75.35	−0.75
Interface AB≡BA	−105.53	−164.59	−59.06	
Total	−2094.43	−2309.49	−215.06	

the self-energy components of the mutated and original protein. It is clearly seen that the self-energy and hence the stability of the mutated protein is improved by −215.1 kcal/mol compared to that of the original protein. One monomer is found to be stabilized by −80.7 kcal/mol and the other is by −75.4 kcal mol. The small differences in values are due to the differences in the initial structure of the two monomers in the crystal structure of the protein. The interface is stabilized by −59.1 kcal/mol. Thus, as it was designed, individually the monomers as well as the interface have been separately stabilized to make the mutated protein really thermostable as a whole.

Fig. 6 shows the cartoon representation of the designed mutated protein with the mutated residues shown in ball and stick along with its mesophilic counterpart on the left. The mutatable (in mesophilic) and the mutated (in the mutated) proteins are separately highlighted. A colored version of this picture is available in the Supplementary Information.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2004.08.011.

3.2. Converting a hetero-dimeric mesophilic protein (1GGP) into a dimeric thermostable protein

It may be mentioned here that it is easier to design a hetero dimer than a homo-dimer as the condition of keeping the two monomers identical is not required in the case of a hetero-dimer.

In the present case the selected protein consists of two different monomers chain-A and chain-B. So, in converting it into a thermostable protein required mutations of the two individual monomers and their interface. The residue-wise interaction pattern of the monomer chain-A is shown in Fig. 7. Based on this pattern we selected the residues Glu-24, Glu-62, Asp-110, Lys-125 and Lys-246 as the key residues as they were polar/charged and interacted rather weakly with the rest of the monomer chain-A. The respective mutatable residues were Val-54, Ala-139, Ala-89, Ala-189 and Ala-203. Side-chain substitution process indicated that the best mutations are Val-54 → Arg-54, Ala-139 → Arg-139, Ala-89 → Arg-89, Ala-189 → Glu-189 and Ala-203 → Glu-203. Fig. 8 represents the $\Delta E_i (= E_i^{Mutated} - E_i^{Meso})$ versus residue number of the monomers chain-A. As in the previous case, here also we find the growth of a number of strong favorable interactions that stabilize the monomer chain-A. For the other monomer chain-B, the residue-wise interaction pattern is represented in Fig. 9. In a similar fashion we selected the residues Lys-62, Arg-63, Asp-122, Glu-189 and Glu-281 as the key residues and Leu-120, Cys-85, Ala-38, Ala-174 and Ile-251 as the mutatable residues. The best mutations for the mutatable residues were found to be Val-54 → Arg-54, Ala-139 → Arg-139, Ala-89 → Arg-89, Ala-189 → Glu-189 and Ala-203 → Glu-203. The values of $\Delta E_i (= E_i^{Mutated} - E_i^{Meso})$ versus residue number of the monomers chain-B are shown in Fig. 10. The growth of several strong interactions is also observed. Thus, the two monomers are stabilized separately through mutations.

Now mutations in the interface were required to stabilize the interface. The residue-wise interaction pattern (Fig. 11) dictates that the residue Glu-180 (chain-A) as the only key residue on chain-A and the residue Cys-169 (chain-B) as the respective mutatable residue on monomer chain-B as shown in Table 2A. Similarly, for the other monomer the key residues were found to be Asp-225 and Arg-137 on chain-B and the respective mutatable residues were found to be Gly-35 and His-234 on chain-A. Side-chain

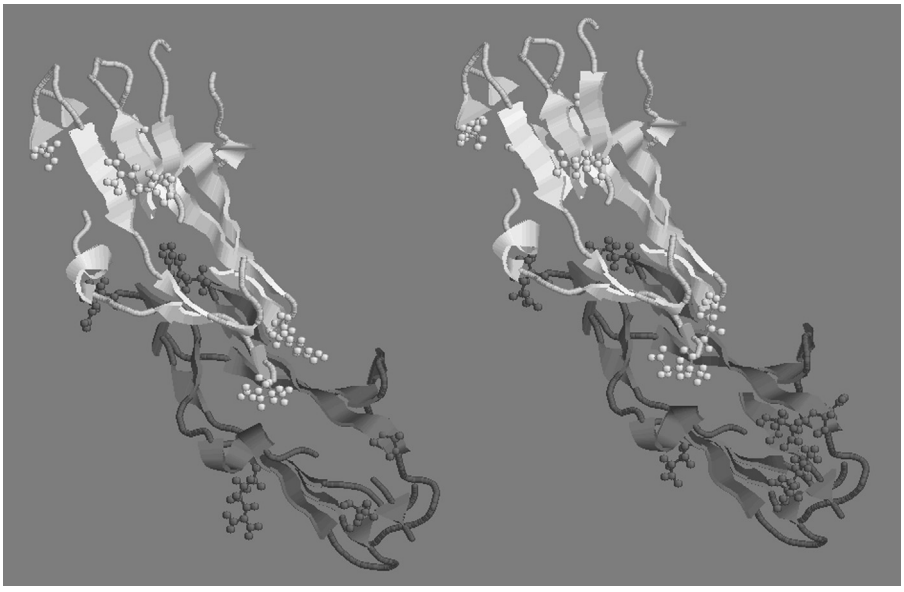

Fig. 6. The cartoon representations of the homo-dimeric mesophilic protein (left picture) and the respective designed mutated thermostable variant (right picture) with the mutatable/mutated residues highlighted in ball and stick. Chain-1 is in black color and the chain-2 is in gray color.

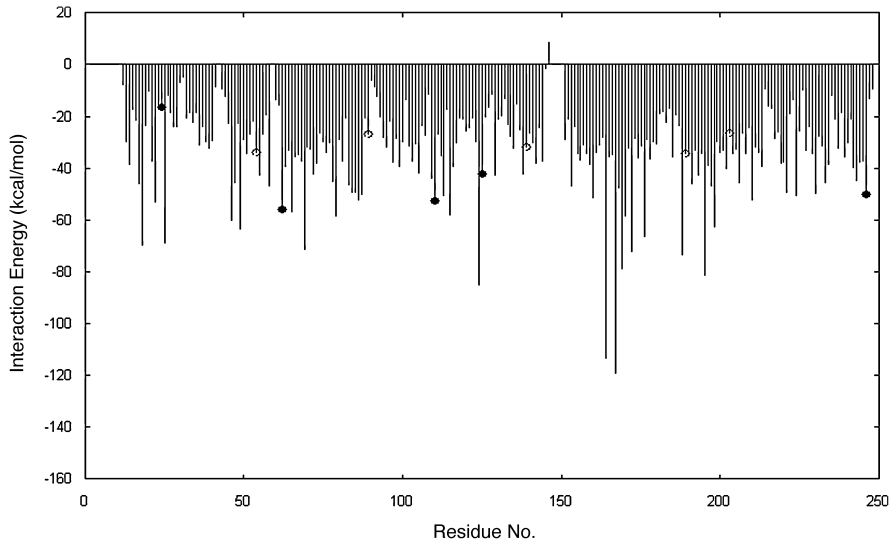

Fig. 7. The pattern of the residue-wise interaction energies of the individual residues of chain-A (monomer-1) of the energy minimized hetero-dimeric mesophilic protein (1GGP) with the rest of monomer-1. The filled circles represent the key residues and the selected mutatable residues are marked by open circles.

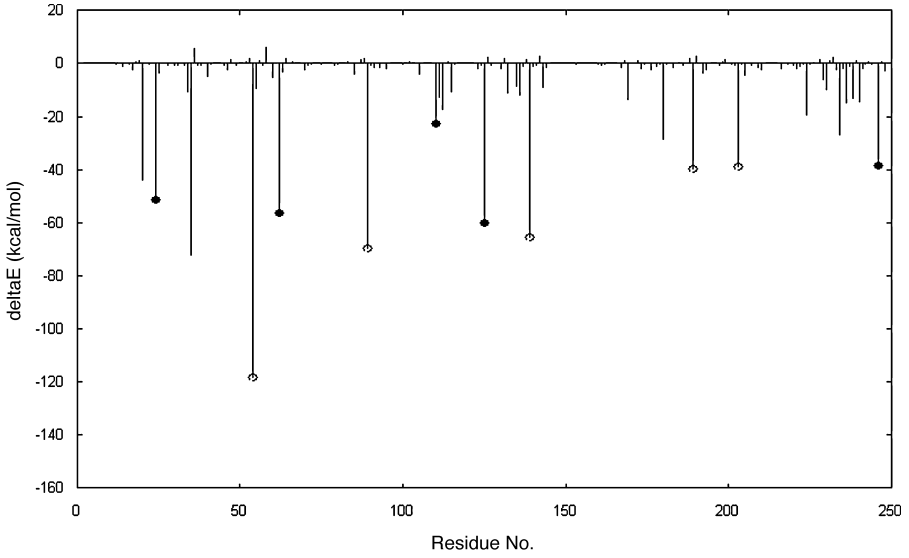

Fig. 8. The pattern of the residue-wise differences $\Delta E_i = E_i^{\text{Mutated}} - E_i^{\text{Meso}}$ in the interaction energies of the individual residues of chain-A between the original protein and the designed mutated protein in the case of the crystal structure 1GGP. The key residues and the mutated residues are highlighted by filled circles and open circles mark the mutatable residues.

Table 2A
Summary of the results on the key residues, mutations, separation of the key residue and the mutatable residue along sequence, gain in energy due to mutation, *etc*., in the case of the protein with PDB ID: 1GGP.

Chain	No.	Key residue	Mutatable residue	Δres	Mutation made	ΔE_i[a] (kcal/mol)
A	1	Glu-24	Val-54	30	Arg-54	−118.46
	2	Glu-62	Ala-139	77	Arg-139	−65.64
	3	Asp-110	Ala-89	21	Arg-89	−69.68
	4	Lys-125	Ala-189	64	Glu-189	−39.54
	5	Lys-246	Ala-203	43	Glu-203	−39.02
B	1	Lys-42	Leu-120	78	Glu-120	−64.82
	2	Arg-63	Cys-85	22	Glu-85	−63.34
	3	Asp-122	Ala-38	94	Arg-38	−73.57
	4	Glu-189	Ala-174	15	Lys-174	−53.52
	5	Glu-241	Ile-251	10	Arg-251	−65.21
Interface AB	1	Glu-180 (chain A)	Cys-169 (chain B)	N.A.	Arg-169 (chain B)	−39.15
Interface BA	1	Asp-225 (chain B)	Gly-35 (chain A)	N.A.	Arg-35 (chain A)	−35.55
	2	Arg-137 (chain B)	His-234 (chain A)		Glu-234 (chain A)	12.10

[a] ΔE_i $(= E_i^{\text{mutated}} - E_i^{\text{meso}})$, where E_i^{mutated} is the interaction energy of the ith residue with the rest of the protein in the designed thermostable variant and E_i^{meso} is its mesophilic counterpart.

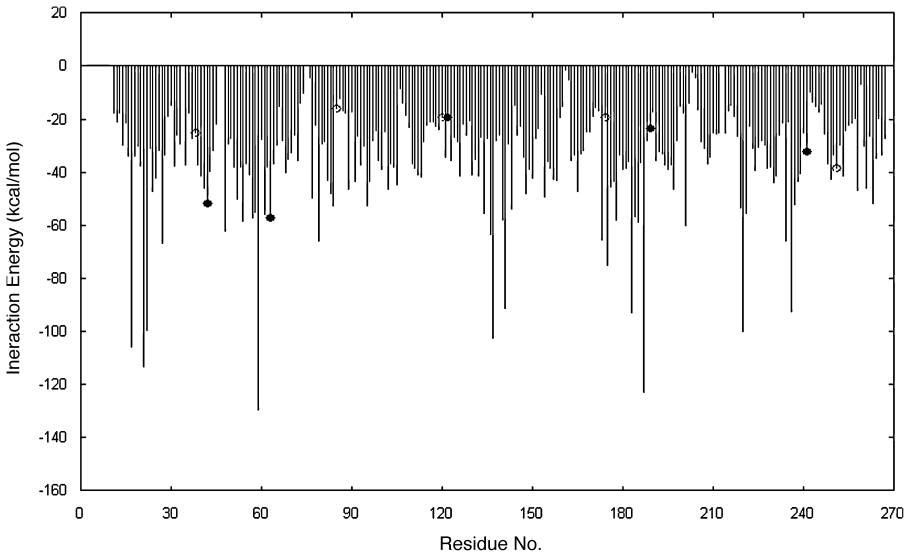

Fig. 9. The pattern of the residue-wise interaction energies of the individual residues of chain-B (monomer-2) of the energy minimized hetero-dimeric mesophilic protein (1GGP) with the rest of monomer-2. The filled circles represent the key residues and the selected mutatable residues are marked by open circles.

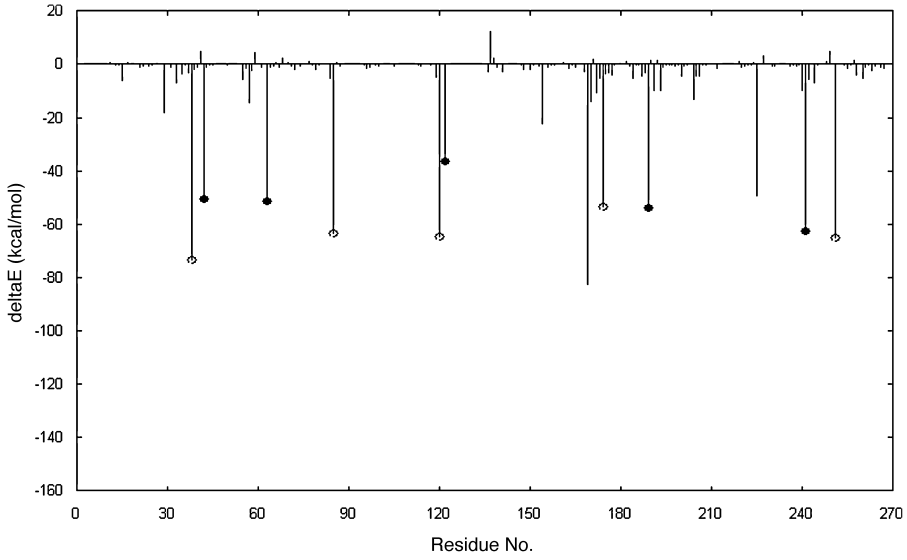

Fig. 10. The pattern of the residue-wise differences $\Delta E_i = E_i^{\text{Mutated}} - E_i^{\text{Meso}}$ in the interaction energies of the individual residues of chain-B with the rest of the monomer 2 between the original protein and the designed mutated protein in the case of the crystal structure 1GGP. The key residues and the mutated residues are highlighted by filled circles and open circles mark the mutatable residues.

substitution process indicated that Cys-169 → Arg-169 on chain-B and Gly-35 → Arg-35 and His-234 → Glu-234 on chain-A are the best mutations in stabilizing the interface. After introducing these mutations by substituting the side-chains by the best side-chains conformers from the libraries we obtained the final mutated protein.

Fig. 12 represents the pattern of the residue-wise differences $\Delta E_i (= E_i^{\text{Mutated}} - E_i^{\text{Meso}})$ in the interaction energies between the original protein and the designed mutated protein at the interface (AB) of the hetero-dimeric protein under consideration. As in the homo-dimer case, here also it is seen that there are a number of large favorable changes in the residue-wise interaction in the interface of the mutated hetero-dimeric protein. Table 2A summarizes the results for this mutated thermostable hetero-dimeric protein. Results show that here the mutated protein was stabilized by −1403.5 kcal/mol. The monomer chain-A was stabilized by −711.6 kcal/mol and the other monomer chain-B was stabilized by −609.4 kcal/mol. The interface was stabilized by −82.5 kcal/mol. The mutation data and the gain in stability are summarized in Tables 2A and 2B.

Fig. 13 shows the cartoon representation of the designed thermophilic protein with the mutated residues shown in ball and stick. The corresponding mesophilic hetero-dimeric protein is also shown on the left. The mutatable (in mesophilic) and the mutated (in the mutated) proteins are separately highlighted. A colored version of this picture is available in the Supplementary Information.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2004.08.011.

4. Discussions

Our original design method in converting a single chain single globule mesophilic protein into a thermophilic protein was validated in our earlier work by demonstrating that several randomly

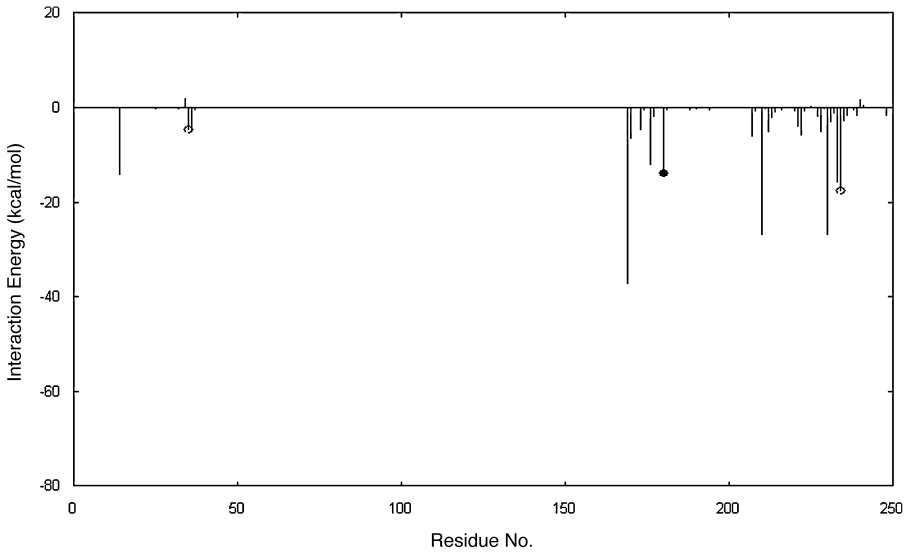

Fig. 11. The pattern of the interaction energies of the residues of monomer-1 (chain-A) with the other monomer (chain-B) in the hetero-dimeric protein (1GGP) against the residue no. of chain-A. The residues with non-zero interaction energy are situated in the interface (AB). The chosen key residues are marked by filled circles and open circles represents the mutatable residues.

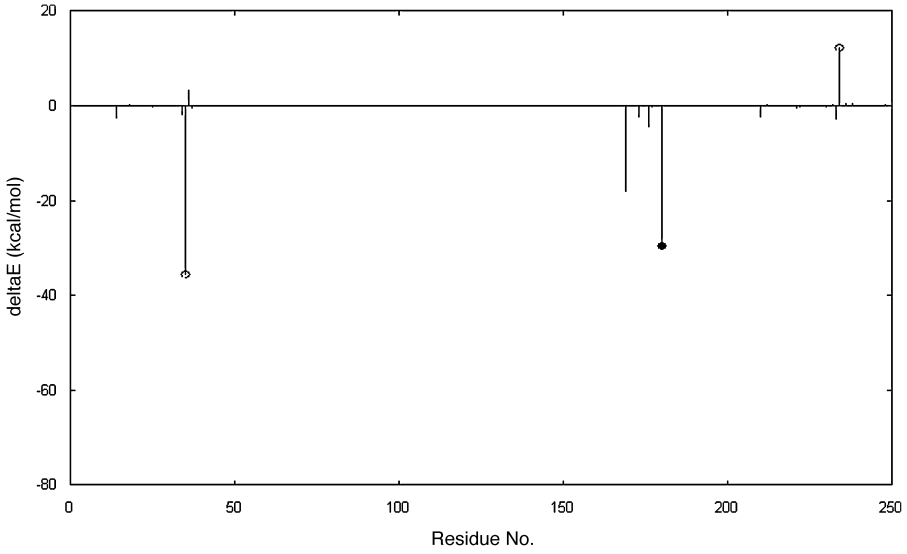

Fig. 12. The pattern of the residue-wise differences $\Delta E_i = E_i^{\text{Mutated}} - E_i^{\text{Meso}}$ in the interaction energies between the original protein and the designed mutated protein at the interface (AB) of the hetero-dimeric protein under consideration. The key residues and the mutated residues are highlighted by filled circles and open circles mark the mutatable residues.

Table 2B
Comparison of the self-energies between the original mesophilic protein (PDB ID: 1GGP) and its designed mutated thermostable variant.

	Self-energy (E^{ms}) of the mesophilic protein (kcal/mol)	Self-energy (E^{mu}) of the designed mutated protein (kcal/mol)	$\Delta E = (E^{\text{mu}} - E^{\text{ms}})$ (kcal/mol)	ΔE/res (kcal/mol/res)
Chain A	−1507.67	−2219.29	−711.62	−3.04
Chain B	−1800.1	−2409.46	−609.36	−2.39
Interface AB≡BA	−257.73	−340.25	−82.51	
Total	−3565.5	−4969.0	−1403.5	

picked natural homologous thermoplilic and mesophilic protein pairs satisfy all the criteria that we used in our approach in converting a mesophilic protein into a thermophilic one [15,16]. In the present work we have extended that approach to cover dimeric proteins as the simplest examples of multimeric proteins. As the experimental verification of this designed protein's thermo stability is beyond our scope, we have used computational measures to show the validity of our approach.

The present manuscript describes and demonstrates how our proposed method of structure-based design of more stable variant of a given protein can easily be extended for dimeric proteins through rational mutations allowing formation of intra-monomer and inter-monomer ion-pairs. We have considered one homo-dimer and a hetero-dimer as examples for demonstration. In a similar way the method can be used to turn a multi-meric mesophilic protein into a multi-meric thermostable protein.

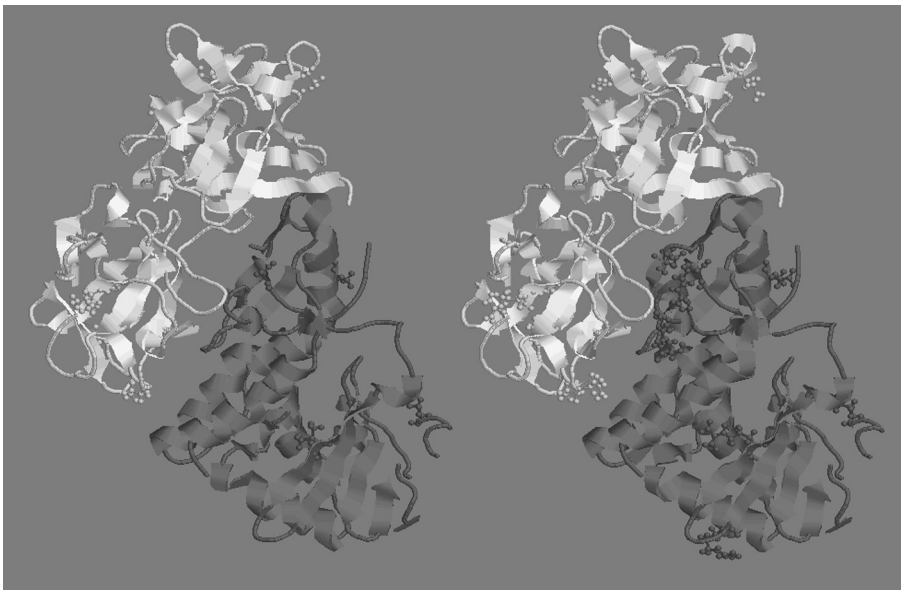

Fig. 13. The cartoon representations of the hetero-dimeric mesophilic protein (left picture) and the respective designed mutated thermostable variant (right picture) with the mutatable/mutated residues highlighted in ball and stick. Chain-1 is in black color and the chain-2 is in gray color.

In the examples, we have only mutated small number of residues and the residues for mutations were chosen at the surface of the protein. Thus, alteration of the fold of the protein by such mutations is quite unlikely. Moreover, we have not selected residues which are known to be related to the local structure making or breaking residues like proline, *etc*.

In order to make the protein strongly thermostable, more mutations may be introduced to the monomers as well as in the interface. However, many mutations in a single monomer may make the protein significantly different from the original one. It may also induce change in the fold. Thus, it is better to introduce minimal number of mutations on both monomers so that each monomer remains very similar to the native one but the surfaces of the monomers and their interfaces are stabilized as desired.

In this work we have used modulation of the electrostatic contribution in the self energy to stabilize a protein. We have computed the electrostatic interaction energies in a rather simple way using Coulomb's law. For a more accurate calculation of the electrostatic energies, one need more sophisticated description of electrostatic interaction. However, the proposed design approach is quite general and the other stabilizing factors can be considered in a similar approach with suitable modifications in the mutation criteria.

It should be mentioned that hydrophobic interaction is known to play an important role in forming dimeric proteins under normal physiological condition. However, it is quite unlikely that hydrophobic interaction will improve the stability of the interface for a thermophilic protein because hydrophobic interaction is weakened with increase in temperatures. Rather, there are evidences that the interfaces are stabilized by ion pairs formed between the monomers [34–36].

5. Conclusions

The present work demonstrates that our proposed method of structure-based design of more stable variant of a given protein can be used for successful conversion of mesophilic homo- and hetero-dimeric proteins into more thermostable ones.

In stabilizing a dimeric protein our method can be applied to stabilize the individual monomers as well as the interface between them by introducing suitable mutations allowing formation of intra-monomer and inter-monomer ion-pairs.

The comparison of the self-energies of the original and the mutated proteins and the interaction energies at the interfaces between the monomers clearly indicates the enhancement of stabilities of the individual monomers and the interface by those mutations.

The comparison of the residue-wise interaction patterns explicitly indicates that the mutated residues are engaged in strong interactions with the rest of the designed mutated protein as originally planned.

References

[1] C. Vieille, G.J. Zeikus, Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability, Microbiology and Molecular Biology Reviews 65 (2001) 1–43.
[2] M. de Champdoré, M. Staiano, M. Rossi, et al., Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest, Journal of the Royal Society, Interface 4 (2007) 183–191.
[3] D.C. Rees, M.W.W. Adams, Hyperthermophiles: taking the heat and loving it, Structure 3 (1995) 251–254.
[4] S. D'Amico, J.C. Marx, C. Gerday, et al., Activity–stability relationships in extremophilic enzymes, Journal of Biological Chemistry 278 (2003) 7891–7896.
[5] R.J.M. Russell, J.M.C. Ferguson, D.W. Hough, et al., The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution, Biochemistry 36 (1997) 9983–9994.
[6] R.A. Criswell, E. Bae, B. Stec, et al., Structures of thermophilic and mesophilic adenylate kinases from the genus *Methanococcus*, Journal of Molecular Biology 330 (2003) 1087–1099.
[7] F. Niehaus, C. Bertoldo, M. Kähler, G. Antranikian, Extremophiles as a source of novel enzymes for industrial application, Applied Microbiology and Biotechnology 51 (1999) 711–729.
[8] H. Atomi, Recent progress towards the application of hyperthermophiles and their enzymes, Current Opinion in Chemical Biology 9 (2005) 166–173.
[9] M.E. Bruins, A.E. Janssen, R.M. Boom, Thermozymes and their applications: a review of recent literature and patents, Applied Biochemistry and Biotechnology 90 (2001) 155–186.
[10] M. Lehmann, C. Loch, A. Middendorf, D. Studer, S.F. Lassen, L. Pasamontes, A.P. van Loon, M. Wyss, The consensus concept for thermostability engineering of proteins: further proof of concept, Protein Engineering 15 (2002) 403–411.
[11] G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins, Journal of Molecular Biology 332 (2003) 449–460.
[12] A. Korkegian, M.E. Black, D. Baker, B.L. Stoddard, Computational thermostabilization of an enzyme, Science 308 (2005) 857–860.
[13] E.S. Zollars, S.A. Marshall, S.L. Mayo, Simple electrostatic model improves designed protein sequences, Protein Science 15 (2006) 2014–2018.
[14] L. Montanucci, P. Fariselli, P.L. Martelli, R. Casadio, Predicting protein thermostability changes from sequence upon multiple mutations, Bioinformatics 24 (2008) i190–i195.

[15] S. Basu, S. Sen, Turning a mesophilic protein into a thermophilic one: a computational approach based on 3D structural features, Journal of Chemical Information and Modeling 49 (2009) 1741–1750.
[16] S. Basu, S. Sen, in: S. Sen, L. Nilsson (Eds.), Thermostable Proteins: Structural Stability and Design, Francis & Taylor, CRC Press, USA, 2011, pp. 117–153 (chapter-6).
[17] G. Vogt, S. Woell, P.J. Argos, Protein thermal stability, hydrogen bonds, and ion pairs, Journal of Molecular Biology 269 (1997) 631–643.
[18] G.R. Grimsley, K.L. Shaw, L.R. Fee, et al., Increasing protein stability by altering long-range coulombic interactions, Protein Science 8 (1999) 1843–1849.
[19] R. Jaenicke, Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? Proceedings of the National Academy of Sciences of the United States of America 97 (2000) 2926–2964.
[20] A. Szilagyi, P. Zavodszky, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comparative survey, Structure 8 (2000) 493–504.
[21] S. Kumar, C.J. Tsai, R. Nussinov, Factors enhancing protein thermostability, Protein Engineering 13 (2000) 179–191.
[22] S. Chakravarty, R. Varadarajan, Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study, Biochemistry 41 (2002) 8152–8161.
[23] A.S. Thomas, A.H. Elcock, Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures, Journal of the American Chemical Society 126 (2004) 2208–2214.
[24] A. Razvi, J.M. Scholtz, Lessons in stability from thermophilic proteins, Protein Science 15 (2006) 1569–1578.
[25] Y. Matsuura, M. Takehira, M. Sawano, K. Ogasahara, T. Tanaka, H. Yamamoto, N. Kunishima, E. Katoh, K. Yutani, Role of charged residues in stabilization of *Pyrococcus horikoshii* CutA1, which has a denaturation temperature of nearly 150 °C, FEBS Journal 279 (2012) 78–90.
[26] A.V. Gribenko, G.I. Makhatadze, Role of the charge–charge interactions in defining stability and halophilicity of the CspB proteins, Journal of Molecular Biology 366 (2007) 842–856.
[27] S. Kumar, B. Ma, C.J. Tsai, et al., Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers, Proteins 38 (2000) 368–383.
[28] N. Kannan, S. Vishveshwara, Aromatic clusters: a determinant of thermal stability of thermophilic proteins, Protein Engineering 13 (2000) 753–761.
[29] D. Perl, F.X. Schmid, Electrostatic stabilization of a thermophilic cold shock protein, Journal of Molecular Biology 313 (2001) 343–357.
[30] B.N. Dominy, H. Minoux, C.L. Brooks 3rd, An electrostatic basis for the stability of thermophilic proteins, Proteins 5 (2004) 128–141.
[31] M. Sadeghi, H. Naderi-Manesh, M. Zarrabi, et al., Effective factors in thermostability of thermophilic proteins, Biophysical Chemistry 119 (2006) 256–270.
[32] R.B. Greaves, J. Warwicker, Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles, BMC Structural Biology 7 (2007) 18–40.
[33] G.I. Makhatadze, V.V. Loladze, D.N. Ermolenko, et al., Contribution of surface salt bridges to protein stability: guidelines for protein engineering, Journal of Molecular Biology 327 (2003) 1135–1148.
[34] A. Letai, E. Fuchs, The importance of intramolecular ion pairing in intermediate filaments, Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 92–96.
[35] F.W. Whipple, E.F. Hou, A. Hochschild, Amino acid–amino acid contacts at the cooperativity interface of the bacteriophage l and P22 repressors, Genes and Development 12 (1998) 2791–2802.
[36] B. Dalhus, M. Saarinen, U.H. Sauer, P. Eklund, K. Johansson, A. Karlsson, S. Ramaswamy, A. Bjørk, B. Synstad, K. Naterstad, R. Sirevag, H. Eklund, Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases, Journal of Molecular Biology 318 (2002) 707–721.
[37] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, et al., CHARMM: a program for macromolecular energy, minmimization, and dynamics calculations, Journal of Computational Chemistry 4 (1983) 187–217.
[38] A.D. MacKerell Jr., et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, Journal of Physical Chemistry B 102 (1998) 3586–3616.