Structural insights into N-terminal to C-terminal interactions and implications for thermostability of a (β/α)$_8$-triosephosphate isomerase barrel enzyme

Pranjal Mahanta[1], Amit Bhardwaj[2,*], Krishan Kumar[2], Vanga S. Reddy[2] and Suryanarayanarao Ramakumar[1]

1 Department of Physics, Indian Institute of Science, Bangalore, India
2 Plant Transformation Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

Correspondence

S. Ramakumar, Department of Physics, Indian Institute of Science, Bangalore-560012, India
Fax: +91-80-2360-2602
Tel: +91-80-2293-2312
E-mail: ramak@physics.iisc.ernet.in

Present address:

Department of Pathology, Kimmel Centre for Biology and Medicine at the Skirball Institute, New York University School of Medicine, New York, NY 10016, USA

Although several factors have been suggested to contribute to thermostability, the stabilization strategies used by proteins are still enigmatic. Studies on a recombinant xylanase from *Bacilllus* sp. NG-27 (RBSX), which has the ubiquitous (β/α)$_8$-triosephosphate isomerase barrel fold, showed that just a single mutation, V1L, although not located in any secondary structural element, markedly enhanced the stability from 70 °C to 75 °C without loss of catalytic activity. Conversely, the V1A mutation at the same position decreased the stability of the enzyme from 70 °C to 68 °C. To gain structural insights into how a single extreme N-terminus mutation can markedly influence the thermostability of the enzyme, we determined the crystal structure of RBSX and the two mutants. On the basis of computational analysis of their crystal structures, including residue interaction networks, we established a link between N-terminal to C-terminal contacts and RBSX thermostability. Our study reveals that augmenting N-terminal to C-terminal noncovalent interactions is associated with enhancement of the stability of the enzyme. In addition, we discuss several lines of evidence supporting a connection between N-terminal to C-terminal noncovalent interactions and protein stability in different proteins. We propose that the strategy of mutations at the termini could be exploited with a view to modulate stability without compromising enzymatic activity, or in general, protein function in diverse folds where N and C termini are in close proximity.

Introduction

Elucidating the molecular basis of protein stability at high temperature continues to attract and fascinate researchers over a broad range of disciplines, and remains a challenge. A number of approaches have been employed to develop stable proteins for biotechnological applications [1,2]. Site-directed mutagenesis together with the comparative analysis of mutant structures is a powerful approach to provide valuable insights into the structural features that govern protein thermostability. Locating the target site of mutagenesis for stability improvement can reduce the screening effort required to find stable mutant(s) by orders of magnitude as compared with random directed-evolution methods [3].

Enzyme stability and activity often appear to trade off at the level of individual mutations. For example, whereas flexibility is required for the catalytic activity of most enzymes, higher thermostability necessitates an increase in the rigidity of the structure. As a result, mutants with increased stability often lose catalytic efficiency [4]. In addition, engineering protein thermostability at the expense of losing enzymatic activity is not biotechnologically desirable. Improving the stability of an already stable enzyme could be advantageous for high-temperature industrial applications. Also, even a modest increase in stability could lead to a > 10-fold longer lifetime [5,6].

Xylanases (EC 3.2.1.8) are glycosyl hydrolases that catalyze the hydrolysis of internal β-1,4-glycosidic bonds of xylan backbones, and have potential economic and environmentally friendly applications in the paper pulp, food, animal feed, and detergent industries, and bio-ethanol and bio-energy production systems [7–10]. A xylanase from *Bacilllus* sp. NG-27 (BSX), which is an extracellular endoxylanase, belonging to glycosyl hydrolase family 10 (GH10), shows optimum activity at a temperature of 70 °C and a pH 8.5 [11]. It has a (β/α)₈-triosephosphate isomerase (TIM) barrel fold, which has been studied concerning its function, structural properties, design, and evolution [12–15]. BSX, apart from thermoalkalophilic features, shows resistance to SDS denaturation and protease K degradation [16]. Hence, BSX serves as an important model system with which to achieve a fundamental understanding of the structure–stability–evolution relationships of the ubiquitous TIM barrel fold.

Earlier studies of recombinant BSX (RBSX) showed that just a single extreme N-terminus mutation, V1L, although not located in any secondary structural element (SSE) (helices or β-sheets), markedly enhanced RBSX stability from 70 °C to 75 °C without loss of catalytic activity [16]. In contrast, the V1A mutation at the same position decreased the stability of the enzyme from 70 °C to 68 °C. However, structural details were not available at that time, precluding any structure-based rationalization of stability changes resulting from a single mutation.

Because of the ring-like architecture of the (β/α)₈-TIM barrel (Fig. 1), the N and C termini come close together in 3D space, enabling several contacts to be formed between the termini. It has been suggested, through mutation studies, that unfolding of the N terminus is one of the first and most critical steps of BSX denaturation at high temperature [16]. In general, terminal residues are exposed to solvent with a low number of nearest neighbor-residues, and are hence considered to have little influence on thermostability [17]. Nevertheless, certain experimental and
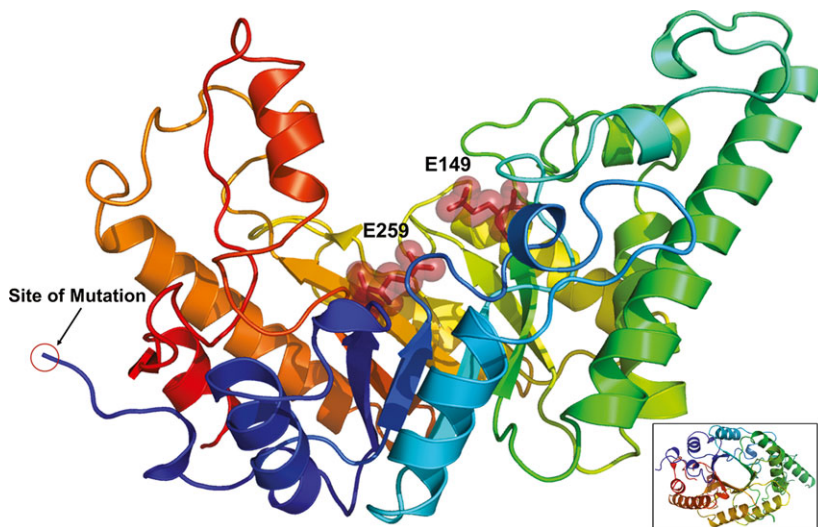
Fig. 1. Overview of the RBSX structure. The structure of RBSX (PDB ID: 4QCE) is shown in cartoon representation (salad-bowl view). Because of the TIM barrel fold, the N terminus (blue region) comes close to the C terminus (red region), and their proximity is implicated in stability enhancement. The mutation is located away from the active site (sphere and stick in firebrick) residues (Glu149 and Glu259). The inset shows the view of RBSX looking down the TIM barrel axis, depicting a ring-like architecture.

computational studies have suggested that protein termini play a role in structural stability and function [18–22]. An *in silico* analysis of a set of two-state folding proteins that took into account the interactions between the terminal SSEs revealed the presence of an N–C motif (N-terminal to C-terminal contacts), and suggested its possible role in initial protein folding, native state stability, and final turnover [23].

Here, we investigated how a seemingly unimportant extreme N-terminus single mutation, although not located in any SSE, affects the structure and interactions, resulting in a change in the thermal stability of the xylanase. To gain structural insights into how the mutation influences the thermostability of the enzyme, we determined the crystal structure of RBSX and the two mutants (V1L and V1A). On the basis of computational analysis, including residue interaction networks (RINs) of the molecular structures of RBSX and its mutants, we demonstrate that augmentation of N-terminal to C-terminal interactions is associated with enhancement of stability of the enzyme RBSX. Perhaps for the first time, we provide a network perspective of the N-terminal to C-terminal interactions, and show that the cumulative effect of a network of noncovalent interactions, which include N-terminal to C-terminal interactions, modulates the thermal stability of the enzyme. The extreme N-terminus mutation is able to induce changes in the structure consistent with protein stabilization, as assessed through contacts and parameters of RINs. Furthermore, we discuss specific examples from different protein families that provide experimental evidence for the connection between protein stability and N-terminal and C-terminal noncovalent interactions. Our observations suggest that mutagenesis at the termini could be exploited to enhance stability without compromising enzymatic activity or function. This may be effective especially in situations where the N and C termini come close together in 3D space, thereby enabling long-range interactions (interactions between residues distantly separated in the primary sequence), as demonstrated here with the example of the TIM barrel fold xylanase, and the same could be extrapolated to diverse folds in both biotechnologically and therapeutically important proteins.

Results

Overall structure of RBSX and comparison with related GH10 xylanase structures

The 3D structure of RBSX adopts the topology of a classical $(\beta/\alpha)_8$-barrel fold (TIM barrel), retaining the characteristic 'salad-bowl' shape observed in GH10 xylanases (Fig. 1). It comprises 12 α-helices, nine β-strands, and five 3_{10}-helices (residues 5–8, 26–30, 94–98, 106–109, and 302–304), as assigned by DSSP [24]. The barrel-forming secondary structures consisting of eight major parallel β-strands lie in the middle, surrounded by eight α-helices (Fig. 1). The C-terminal ends of the β-strands of the β-barrel form a long and open cleft on the protein surface. The active site cleft is exposed to the solvent, and corresponds to the catalytic/activity face of the enzyme [25].

RBSX with an additional Met resulting from the start codon belongs to GH10. RBSX differs from many other xylanases in being operational under alkaline condition (pH 8.5) and at elevated temperature (70 °C). Several GH10 xylanase structures have been solved to date (http://www.cazy.org/GH10.html). However, only one report is available on an alkaline-active (pH 10) thermostable (70 °C) GH10 xylanase structure [Protein Data Bank (PDB) ID: 2UWF] produced by an alkalophilic organism, *Bacillus halodurans* S7 [26]. *B. halodurans* S7 xylanase shows 78% sequence identity with RBSX. Structures are available for a few other GH10 xylanases, which are produced by nonalkalophilic organisms and do not show significant operational stability at higher pH [11,26]. The desirable properties of xylanases in the paper and pulp industries are stability and activity at high temperature and alkaline pH [27]. It is reported that alkaline xylanases have more surface-accessible acidic residues than their nonalkaline counterparts [11,26]. The adaptation to extreme conditions such as alkalophilicity can take place through insertion or exchanges of short sequences without the requirement for gradual changes over the entire chain [11].

Comparative analysis of structures of the enzymes with stabilizing and destabilizing mutations

In an earlier study from our group, different extreme N-terminus mutants of RBSX were generated in which the first residue, Val1, was replaced with various amino acids, and each mutant was subsequently analyzed for its thermal stability at high temperatures. The thermal stabilities of RBSX and the mutant proteins were determined by CD measurements at a range of temperatures [16]. RBSX and its mutants were found to unfold in an irreversible manner, and the apparent melting temperature (T_m) was calculated for all protein samples with a constant temperature slope of 60 °C·h^{-1}. It was observed that the V1G and V1A mutations decreased the stability of the protein by

12 °C and 2 °C, respectively, whereas the V1F and V1D mutations did not significantly change its thermostability, as compared with RBSX [16]. On the other hand, the V1L mutation markedly enhanced the thermostability of RBSX from 70 °C to 75 °C without compromising its catalytic activity, and resulted in higher cooperativity in the thermal unfolding transition [16]. However, at present, structural details are not available for any mutants except the V1A and V1L mutants. A brief summary of crystallization, data collection, structure solutions and refinement statistics of RBSX and the V1A and V1L mutants is given in Table 1 and Doc. S1. Comparison of the molecular

Table 1. X-ray data collection and refinement statistics.

Name	RBSX	V1A	V1L
Crystallographic data			
Space group	$P2_12_12_1$	C2	$P2_12_12_1$
Unit cell dimensions			
a (Å)	54.77	73.57	54.88
b (Å)	75.65	80.12	76.58
c (Å)	176.91	69.90	176.73
α (°)	90	90	90
β (°)	90	110.81	90
γ (°)	90	90	90
Unit cell volume (Å³)	733 000	385 141	742 745
Data collection			
Temperature (K)	100	100	100
Resolution (Å)	27.7–2.32	26.8–2.26	40.2–1.96
	(2.45–2.32)	(2.38–2.26)	(2.07–1.96)
Total reflections	30 910	17 095	52 480
Unique reflections			
Above 1σ	30 459	17 093	52 436
Above 3σ	21 329	15 556	45 630
R_{merge} (%)	16.1 (33.8)	3.9 (15.3)	7.6 (21.3)
Average $I/\sigma(I)$	6.7 (3.8)	25.5 (10.0)	18.5 (8.7)
Completeness	95.1 (96.3)	96.3 (94.5)	97.3 (94.3)
Redundancy	4.7 (4.5)	5.2 (5.1)	6.9 (6.6)
Solvent content (%)	44.97	47.55	45.14
Refinement statistics			
R_{work}/R_{free}	17.9/22.7	17.4/22.5	15.2/19.0
No. of atoms			
Protein	5807	2905	5854
Ligand/ion	4	5	16
Water	419	202	434
Average B-factors (Å²)			
Protein	21.6	26.3	13.9
Ligand/ion	15.6	26.4	18.74
Water	20.6	27.2	23.2
rmsd			
Bond distance (Å)	0.007	0.005	0.019
Bond angles (°)	1.123	0.955	1.802
Luzzati coordinate error (Å), working set	0.279	0.269	0.175
PDB entry	4QCE	4QCF	4QDM

structures of the V1A and V1L mutants with that of RBSX showed no significant changes in the overall 3D structure of proteins, despite the difference in their thermal stabilities. The overall C_α rmsd between RBSX and the V1L mutant is 0.393 Å, whereas that between RBSX and the V1A mutant is 0.265 Å. The question therefore arises of what may be the mechanism of thermal stabilization/destabilization, considering that there is only a minimal change in their overall 3D structures. The mutation is located on an extended loop at the extreme N-terminal region, and makes no dominant interaction exclusively by this first residue (Leu1 in the V1L mutant and Val1 in RBSX). This observation raises the possibility of effects of a noncovalent interaction network that transmits changes near and far from the site of mutation, and changes the overall stability of RBSX. In a folded protein, a network of interactions brings the distal residues in a sequence space to close proximity in 3D space. This extensive network of interactions gives proteins structural flexibility, integrity, and thermal stability [28,29]. Therefore, we focused on residue contacts and RINs to investigate, from a network perspective, this cumulative nature of thermal stabilization–destabilization and to identify the changes in both local and nonlocal interactions between the structures. The aim of the present study was not to obtain the most stable structure by carrying out all possible mutations at the extreme N terminus, but to gain structural insights into the modulation of stability caused by mutations in the terminal region.

Changes in contacts in the vicinity of the mutation

Although there was an increase of ~ 5 °C in the thermostability resulting from a single Leu mutation, analysis of the interactions undergone by the mutated residue and recombinant protein showed that Leu1 in the V1L mutant structure undergoes similar types of interaction (van der Waals) as Val1 in the RBSX structure. However, because of its greater bulk and better conformational accessibility than Val1 in RBSX, the Leu side chain in the V1L mutant structure forms more van der Waals contacts with side chain atoms of Arg344 (C-terminal residue) than the Ala side chain in the V1A mutant structure (Fig. 2; Table S1). Furthermore, these additional cohesive contacts made by Leu1 influenced the relative decrease in the solvent-accessible surface area of Arg344 by 19.6% in the V1L mutant in comparison with Arg344 in RBSX. The decrease in the solvent-accessible surface area of Arg344 is even more pronounced (29.8%) when the structures of the
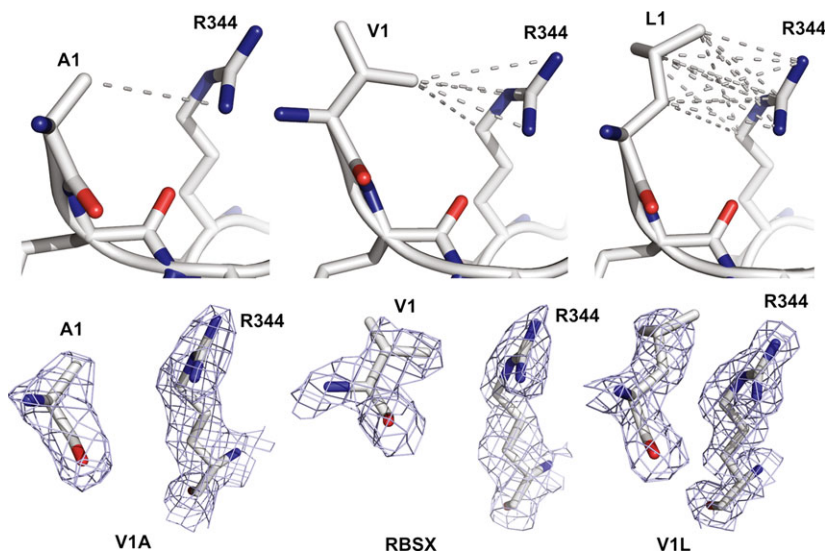
Fig. 2. Overlay of van der Waals contacts at the site of mutation. Van der Waals contacts between the side chain atoms of the mutated residue and surrounding side chain atoms are shown. The dashed line represents the van der Waals contact at a distance cut-off of 5 Å. The lower panel shows the $2F_o - F_c$ electron density map of corresponding residues, contoured at the 1.0_σ level with side chains as sticks.

V1L and V1A mutants are compared. Evaluation of atomic packing according to the small-probe contact dot surface between the mutated residue and Arg344 with the RINERATOR module [30] showed a higher noncovalent interaction score (I_s) between Leu1 and Arg344 (2.66) for the V1L mutant structure than between Val1 and Arg344 (0.063) for the RBSX structure, and between Ala1 and Arg344 (0.00) for the V1A mutant structure. Here, the noncovalent I_s for the interacting residues is suggested to be proportional to the strength of the interaction (the higher the score, stronger the interactions of the connecting residues), which is the weighted sum of the nonoverlapping van der Waals contacts, the volume of hydrogen bonds, and atomic overlaps (clashes) [30,31]. As the mutated residue belongs to an extended loop in the extreme N terminus of the protein, and its interacting partner Arg344 belongs to the C-terminal region of the protein structure, this result draws attention to the role of the N-terminal and C-terminal regions in contributing to the overall stability through mutual interactions.

Analysis of residue-wise C_β contact order (RWC$_\beta$CO)

As discussed, side chain atoms of Leu1 are in close proximity to side chain atoms of Arg344 in the folded 3D structure. Despite the introduction of such a bulky residue (Leu1), we observed that the distance between the C_β atoms of Leu1 and Arg344 in the V1L mutant structure is smaller (6.48 Å) than that between the corresponding residues in the RBSX structure (7.12 Å) and the V1A mutant structure (7.21 Å). Thus, the distance between C_β atoms belonging to interacting

residues is shorter in the more stable mutant than in the less stable mutant, suggesting that the distances between C_β atoms could be used to gain information about the mutation-induced structural change. Hence, we evaluated RWC$_\beta$CO [32] (Doc. S1) to investigate the influence of long-range interactions in the protein structures. The analysis of RWC$_\beta$CO for all three structures (V1L, RBSX, and V1A) showed that higher values of RWC$_\beta$CO belong to the termini of the protein (Fig. S1). This is because of the occurrence of a number of contacts between the N-terminal and C-terminal regions, as they are in close proximity in the 3D structure. We then computed ΔRWC$_\beta$CO (RWC$_\beta$CO$_L$ – RWC$_\beta$CO$_A$), the difference in RWC$_\beta$CO value between the more stable mutant (V1L) and the less stable mutant (V1A), to compare the changes in the protein structure resulting from mutation, and plotted it along the polypeptide chain. It was observed that there is a significant increase in ΔRWC$_\beta$CO in the N-terminal and C-terminal regions, although small differences are present in other parts of the protein structure (Fig. 3). This is one piece of evidence that both terminal regions are substantially affected by the mutation in comparison with the other parts of the protein structure, and that it could be playing an important role in contributing to the overall stability through the enhancement of long-range interactions.

Analysis of noncovalent interactions

To assess the contribution of chain termini to RWC$_\beta$CO values and their relationship with mutant stability, we studied the differences in atom–atom contacts at the terminal regions in detail. We used a
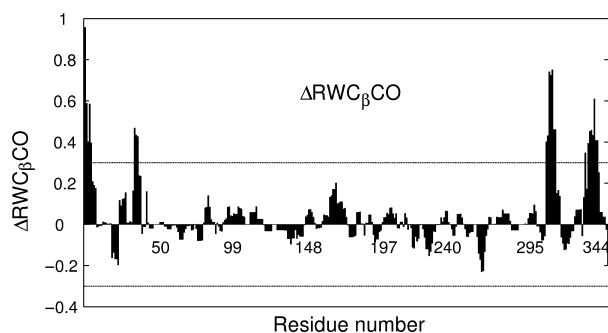
Fig. 3. Difference of $RWC_\beta CO$ between the more stable and less stable mutants: five-residue moving average of the difference in $RWC_\beta CO$ between more stable (V1L) and less stable (V1A) mutant structures. The dashed line corresponds to the value of $\Delta RWC\beta CO$, which has a *Z*-score of 2 (Doc. S1).

distance cut-off of 5 Å (the higher cut-off for attractive London–van der Waals forces [33]) to capture only effective physical contacts within and between the terminal atoms. Examination of atom–atom contacts between an N-terminal segment [residues 1–25; up to the second SSE from the N-terminal end] and a C-terminal segment (residues 319–354; up to the second SSE from the C-terminal end) for each structure revealed that the V1L mutant has a higher value of normalized atom–atom contacts between termini than the V1A mutant (Table S2; Fig. S2). We observed that there is an increase of ~ 11.3% in the contacts between the N-terminal segment and C-terminal segment in the V1L mutant as compared with the V1A mutant. Also, there is an increase of ~ 3.4% in the contacts within the N-terminal segment in the V1L mutant as compared with the V1A mutant. Thus, the increase in contacts is more pronounced in the N-C-terminal region than within the N-terminal region, indicating a 'cosying up' of the terminal regions i.e. enhanced mutual interactions between the termini, in the more stable mutant. Furthermore, we observed that the noncovalent I_s between the N-terminal segment and C-terminal segment as assessed with the RINERATOR module [30] is higher in the V1L mutant structure

than in the V1A mutant structure (103.5/73.5). These observations collectively suggest that the substantial increase in thermostability of the V1L mutant could be a result of better relative contributions of various types of interaction, particularly those between terminal regions.

A network perspective of protein stability

In protein structures, there are a variety of weak and strong noncovalent interactions that integrate different parts of the structure. It is the interplay of these interactions that provides structural stability. We analyzed this by using a network representation of protein structure by generating a RIN, which considers all noncovalent interactions between pairs of interacting residues (Experimental procedures). We also decomposed the network into different subnetworks based on the strength of intercation (noncovalent I_s) between interacting residues, and analyzed their global topology and network parameters. We obtained well-known network parameters, such as total number of edges or links (*E*), the edge/node ratio (*E*/*N*, where *N* is the total number of residues in the protein structure), and average number of nearest neighbors (*<n>*), for the more stable mutant (V1L) and the less stable mutant (V1A) at different I_s cut-off (I_{smin}) values (Table 2). A lower *E*/*N* ratio of a network indicates that most nodes are isolated and that the graph is largely disconnected. As the *E*/*N* ratio increases, connectivity between the nodes in the graph increases. On the other hand, a higher *<n>* value indicates higher average connectivity of a node in the network. We observed that the values of these network parameters are very similar to each other in all three structures. This is understandable, because all three structures have the same size (354 residues), and only one residue is mutated in their primary structures. However, it is relevant to compare their network parameters, as they show different thermostability scales. The analysis revealed that all three network parameters (*E*, *E*/*N*, and *<n>*) have higher values for the V1L mutant than for the V1A

Table 2. Different network parameters of RBSX and its mutant structures at different I_{smin} values.

	I_{smin}								
	0			1			3		
	E	*E/N*	*<n>*	*E*	*E/N*	*<n>*	*E*	*E/N*	*<n>*
V1A	4587	12.96	8.53	1831	5.17	4.95	341	0.96	1.85
RBSX	4594	12.98	8.56	1867	5.27	4.99	363	1.02	1.90
V1L	4699	13.27	8.62	1895	5.35	5.05	363	1.02	1.98

mutant at different I_{smin} values. The greater number of interactions (higher E and E/N values) and better connectivity within the network ($<n>$) at different I_{smin} values is likely to influence structural stability, providing the extra stabilization for the V1L mutant in comparison with RBSX or the V1A mutant [28]. We can infer that the observed difference in network parameters is a result of the combined effect of various subtle changes in interactions (hydrogen bonds, van der Waals, ion pairs, etc.) manifested throughout the structure because of a single mutation.

The size of the largest cluster in the network analysis is often used to understand the nature and connectivity of the network [28,34]. Here, we compared the change in size of the largest cluster as a function of I_{smin} for both mutant structures. We observed that, irrespective of the difference in thermostability scales, the normalized size of the largest strongly connected component (LSCC) (in terms of nodes or number of residues) gradually decreases with an increase in I_{smin} in both mutants. The normalized size of the LSCC undergoes a sharp transition after a particular I_{smin} cut-off, which begins around $I_{smin} = 1$ and lies within a narrow range of I_{smin} (1–2), with no major change towards the side of higher interaction cut-off (Fig. 4). A similar trend has been observed in other studies [28,34]. This transition in the LSCC is attributed to the loss of different noncovalent interactions in the networks as I_{smin} increases, thus quickly generating a large number of small clusters. It is of note that the more stable mutant (V1L) not only has higher E values and higher E/N ratios for different

I_{smin} values than the V1A mutant, but also has larger size of the LSCC in their network (Fig. 4). The cooperative nature of different stabilizing interactions in the network seems to be positively influencing other interactions, as shown by the sizes of the LSCC at different I_{smin} values in the more stable mutant structure. Our result is consistent with the previous findings that thermostable proteins have improved structural attributes, such as better connectivity and higher number of nodes, at different I_{smin} values than their corresponding mesophilic homologs [28,34].

Comparison of RINs between the V1L and V1A mutants and the importance of terminal interactions

To explore the difference between the mutant structures, we constructed a comparison network of the LSCCs of the V1L and V1A mutants at $I_{smin} = 1$ by using the RINALYZER plugin [30] of CYTOSCAPE [35]. In this comparison network, each node represents a pair of aligned residues, and an edge exists if there is a noncovalent interaction between the connecting nodes in either of the two compared RINs [31]. The combined comparison network of the V1L and V1A mutants generated on the basis of the superposition alignment of the corresponding 3D structures resulted in 535 identical edges that correspond to noncovalent interactions for both the V1L mutant and the V1A mutant. We found that the more stable mutant (V1L) has a considerably higher number of nonidentical noncovalent interactions, i.e. 192 edges, which correspond to 28 unique residues (nodes), than the less stable mutant (V1A), i.e. 125 edges, which correspond to 14 unique residues (nodes) respectively (Table S3). This comparison network provides further information about the location of unique residues that have the largest change in local residue interactions in the protein structure. It is notable that many of these unique residues are distributed in and around the terminal regions of the protein (Fig. 5; Table S3). We observed that ~ 25% of these unique residues in the V1L mutant and ~ 14% of unique residues in the V1A mutant correspond to the N-terminal region and the C-terminal region of protein structures, implying the importance of interactions involving terminal residues (within termini and/or between termini) that might have an impact on their stability. These differences in the numbers of unique noncovalent interactions (nonidentical edges) and residues (number of unique nodes) in the comparison network indicate that there is a perturbation in the RINs brought about by a single extreme N-terminus mutation.
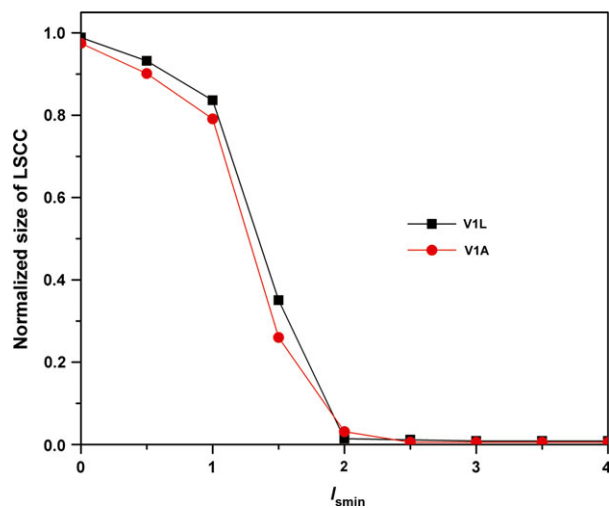

Fig. 4. Transition profile of LSSCs in the network: differences in transition profile in the size of LSCCs of the V1L and V1A mutant structures at different I_{smin} values.
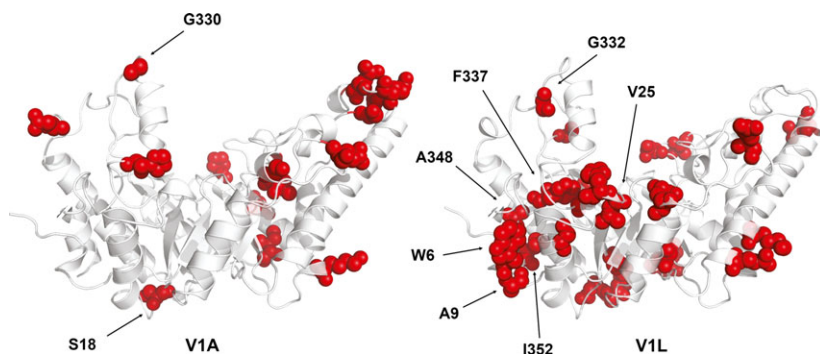
Fig. 5. Comparison of RINs of the LSSC at $I_{smin} = 1$. The positions of unique residues (red spheres) in the 3D structure of the V1L and V1A mutants are shown. The unique residues belong to the cluster of the LSCC when compared between the RINs of the V1L and V1A mutants at $I_{smin} = 1$. Arrows point to terminal residues.

Discussion

The present findings provide valuable insights into the role of direct noncovalent interactions between the N and C termini in protein stabilization. Figure 2 provides an example of such interactions that are enhanced in the more stable mutant. The direct N-terminal to C-terminal contacts in the V1L mutant involving Leu1 show a clear difference in the degree of packing interactions of the side chain atoms in comparison with Val1 in RBSX and Ala1 in the V1A mutant (Fig. 2; Table S1). It appears that these additional interactions might play a role in tying down the extreme N terminus during thermal unfolding at high temperature. Furthermore, we observed an enhancement in the number of overall N-terminal to C-terminal direct contacts in the more stable mutant structure (V1L), whereas the absence of many N-terminal to C-terminal contacts could increase local unfolding of the peptide chain at these weak links, and result in a lower unfolding temperature for the V1A mutant (Fig. S2; Table S2). Fraying of the terminal regions may make a protein susceptible to unfolding at high temperature. Hence, it may be advantageous if the terminal regions dock with each other and mutually stabilize, thereby reducing susceptibility to unfolding at high temperature [23]. However, one cannot neglect the interactions between the terminal residues and other parts of the protein structure, and their independent role in protein stabilization [18,19,22].

The substitution of residues with different physicochemical properties and their location in the tertiary structure can cause changes in residue–residue contacts [36]. In order for the structure to adapt to the substitution of the new side chain, there is often a rearrangement of native contacts by the neighboring residues. The extent of this rearrangement of residue contacts depends on how connected the region of substitution is with the rest of the structure. We observed a structural rearrangement of contacts throughout the structure, and more so within and between terminal regions. The cooperative nature of these stabilizing contacts indirectly or allosterically propagates to the other parts of the structure, and positively influences other interactions, as shown by network analysis, where the LSCC is larger for all I_{smin} values in the more stable mutant structure (V1L) than in the less stable one (V1A) (Fig. 4). Thus, it is likely that the increased stability shown by the V1L mutant results from cumulative effects of small changes rather than solely the effect of interactions involving the substituent residue. This effect is reminiscent of the concept in economics of 'comedy of the commons' in property resources [37], applied here to protein stabilization, in which a cumulative effect of many contributions leads to a desired outcome, in this case protein stability. In addition, residues at long distances from each other in the primary structure play an important role in stability of the protein, as shown by the analysis of RWC$_\beta$CO and RINs (Fig. 3; Table S3). Obviously, N-terminal to C-terminal contacts are the longest-range interactions possible in terms of sequence separation in any given protein. These results suggest that the overall increase in long-range interactions (primarily through N-terminal to C-terminal contacts) in the V1L structure upon mutation is one of the primary sources of the increase in thermal stability. Our results are consistent with earlier findings that long-range interactions, connecting different parts of the protein structure, have a major role in folding and stabilizing the tertiary structure of the protein [34,38,39]. However, what is remarkable is that all of these structural changes are elicited by just a single mutation not located in any SSE, at the extreme N-terminus of the protein.

A few other studies have shown the role of interactions between termini in modulating the stability of proteins. In a GH10 xylanase from *Aspergillus niger*, deletion of terminal disordered residues reinforced the contacts between the N and C termini, providing additional compactness to the structure, and thereby

increasing protein stability [20]. In a recent mutational analysis of a similar xylanase, Song et al. [22] suggested that strengthening the hydrophobic interactions within the N-terminal element and between the N and C termini is responsible for the improved stability of the enzyme. A study of chimeric proteins by exchanging the N or C terminus from a thermophilic TmxAcat xylanase and a hyperthermophilic TmxB family xylanase from *Thermotoga maritima* MSB8 showed that interactions between the N and C termini contribute significantly to thermostability [40]. In an earlier study, we provided experimental data showing that the interaction between the N-terminal and C-terminal regions formed by Phe4, Trp6 and Tyr343 (aromatic cluster) is an important determinant of RBSX stability and folding [41]. In addition, information gained from deep sequencing of the TEM-1 β-lactamase enzyme has revealed that several terminal residues are sensitive to substitutions, suggesting a possible role of these residues in enzyme stability, solubility, or catalytic activity [42]. However, crystal structures are unavailable either for the deletion mutants or for the substitution mutants, which could help in providing a structure-based rationalization of stability changes resulting from interactions between termini.

The importance of N-terminal to C-terminal contacts in protein stability has also been reported for other proteins. One such example comes from the homologous pairs of cold shock proteins from the mesophilic *Bacillus subtilis* (Bs-CspB, T_m = 53.9 °C) and the thermophilic *Bacillus caldolyticus* (Bc-Csp, T_m = 76.9 °C). Both Bs-CspB and Bc-Csp are small,

monomeric proteins composed of 67 and 66 residues, respectively; they do not contain any disulfide linkages, and they differ in sequence by only 12 residues [43]. It was found that the additional stability of Bc-Csp largely originates from the contribution of Arg3 (N terminus) and Leu66 (C terminus). Furthermore, mutational analysis revealed that substitution of Arg3 by Glu (the equivalent residue in Bs-Csp) decreased the stability of Bc-Csp by ~ 4 kJ·mol^{-1}. Structural analysis of Bc-Csp and the R3E mutant showed that van der Waals interactions between Arg3 and Leu66 in Bc-Csp became more extensive in comparison with Glu3 and Leu66 in the R3E mutant. The authors pointed out that the decreased number of overall N-terminal to C-terminal van der Waals interactions resulting from the shorter side chain of Glu is responsible for the loss of stability in Bc-Csp [43,44]. Interestingly, by analyzing the precomputed RIN from the RINdata web service (http://rinalyzer.de/rindata.php), we found that Arg3 in Bc-Csp (PDB ID: 1C9O) has a better noncovalent I_s (19.64) than Glu3 (7.03) in the R3E mutant crystal structure (PDB ID: 1I5F). There have been other experimental reports of the effects of terminal mutations on the native state stability of proteins (Fig. 6). For example, a cavity-creating mutant at the C-terminal end (I96A, I_s = 19.3/10.8) of barnase destabilized the structure by 3.52–4 kcal·mol^{-1} [45], whereas mutation of another terminal residue (V2T; I_s = 5.3/4.7) of ribonuclease from *Streptomyces aureofaciens* destabilized it by ~ 0.9 kcal·mol^{-1} [46]. We also observed that, in both mutant structures, there is a reduction in N-terminal to C-terminal interactions by
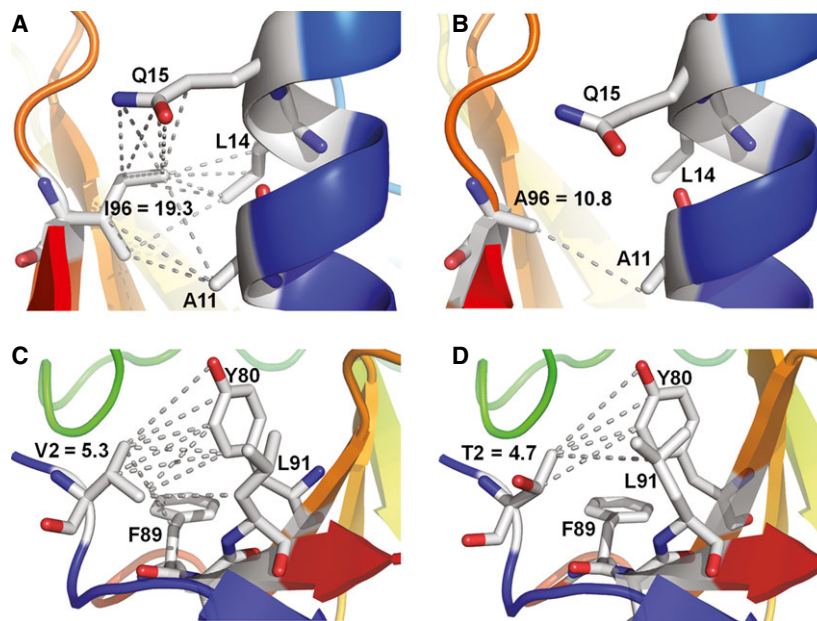
Fig. 6. The association of N-terminal to C-terminal contacts with protein stability. Van der Waals contacts made by the mutated residue with the residue(s) belonging to the terminal region are shown. (A, B) The native and mutated (I96A) structures of barnase (PDB ID: 1BNI and 1BRK), respectively. (C, D) The native and mutated (V2T) structures of ribonuclease from *S. aureofaciens* (PDB ID: 1RGG and 1UCI respectively. The numerical values represent the noncovalent I_s as assessed by the RINERATOR module [31], and are proportional to the strength of interactions between connecting residues. For both proteins, the N terminus (blue) comes into close proximity to the C terminus (red) in the 3D structure. Here, PDB IDs 1BNI and 1RGG correspond to more stable proteins.

the mutated residues, indicating the possible importance of interactions between termini in modulating their stability (Fig. 6).

Many proteins have their N and C termini in contact with each other, although no physical principle has been stated regarding why they do so. An *in silico* analysis on structures of single-domain proteins taking into account the interactions between the terminal SSEs showed the presence of N–C motifs (N-terminal to C-terminal contacts), and suggested a possible role in initial protein folding and stability [23]. One computational analysis that addressed the role of N-terminal to C-terminal coupling in the folding transition of single-domain protein from three different protein folds suggested that switching off the N-terminal to C-terminal interactions decreases folding cooperativity and substantially lowers the free energy barrier of the folding transition state [47]. Therefore, fortifying the interactions between the terminal regions might help to increase the initial activation energy barrier, resulting in enhanced resistance against global unfolding at higher temperature. The present study reiterates as well as expands on those findings, as the mutated residue that modulates (stabilizes/destabilizes) the stability of RBSX through reinforcement of N-terminal to C-terminal interactions belongs to the extreme N terminus of the protein, and is not part of any SSE. Furthermore, our study also provides a network perspective of the interactions involving terminal residues, showing that changes are not restricted to the terminal regions, and propagate to other parts of the protein structure (Fig. 5; Table S3).

Our work suggests that augmenting N-terminal to C-terminal noncovalent interactions is associated with an enhancement of protein stability. Such stabilization presumably protects against unfolding of an already folded protein, and may aid the folding process [23]. Although it is clearly possible to stabilize proteins with other mechanisms/factors, as previously reported [48,49], we have demonstrated that proteins can be stabilized without compromising their biological functions through optimization of N-terminal to C-terminal noncovalent interactions. Because of the ring-like architecture of the $(\beta/\alpha)_8$-TIM barrel (Fig. 1), the N and C termini come into close proximity in 3D space, providing opportunities for stabilization through mutual interactions. This apparent stabilization through N-terminal to C-terminal interactions could be seen to be implicated in the structures of TIM ([EC 5.3.1.1](#)) and the NAD(P)-binding Rossmann-fold domain protein glyceraldehyde-3-phosphate dehydrogenase (an example of a non-TIM barrel fold protein) isolated from different organisms when

compared across monomers, despite their similar 3D structures (Doc. S1; Fig. S3; Table S4). In our study, based on the comparative analysis of crystal structures, we explicitly show that the more stable mutant is associated with better N-terminal to C-terminal interactions. When taken together, these observations support the connection between N-terminal to C-terminal noncovalent interactions and protein stability.

Conclusions

Sequence and structure-based bioinformatics analyses have delineated a methodology to identify target positions for mutagenesis that would enhance protein thermostability. In this context, our study suggests that protein termini constitute one of the regions of interest for designing an effective mutagenesis library (reduced size) in a targeted way, with a view to improving protein stability, thus adding to the repertoire of approaches for increasing the thermal stability of proteins and giving our results wider applicability. It seems interesting that a number of important folds and superfolds [50] have their N and C termini in contact with each other, thus offering opportunities for modulating stability through mutual interactions. 'Making the two ends meet' seems to be a feature common to all of these proteins. It is tempting to speculate that proteins might have evolved the N-terminal and C-terminal interactions as one of the strategies to stabilize their structures in a fold-specific manner. Because, in diverse folds/proteins, the terminal regions are in close proximity, it may be suggested that they could be considered as candidates for modulating stability by mutation focusing on terminal regions, in contrast to the general belief that terminal residues are very flexible and might have less effect on stability [17]. It is important to investigate more proteins from diverse organisms to decipher other biologically significant aspects of N-terminal to C-terminal contacts. Eventually, such studies should help in understanding the evolution and utilization of interactions between termini in the protein universe, and in developing effective protein engineering strategies.

Experimental procedures

RINs

Protein structure can be represented as a RIN between residues [28,29]. In this analysis, we used the RINERATOR module [30] to generate a RIN from the 3D structure of each protein. A RIN consists of nodes that represent residues, and are connected by edges or links that correspond

to noncovalent interaction between the nodes (residues). RINERATOR creates an undirected weighted network with multiple edges, in which the edges are defined on the basis of the noncovalent I_s between the connecting nodes. RINs are generated from a protein structure with the following steps. All hydrogen atoms are added to the original protein structure by the REDUCE [51] program, PROBE [52] is then used to identify noncovalent interactions between two interacting residues, and a RIN is then generated by the RINERATOR package. The edges in the RIN are labeled with different interaction types, e.g. interatomic contacts, hydrogen bonds, overlapping van der Waals radii, and generic residue interactions. The edges are weighted with the respective noncovalent I_s for the interacting residues as computed by PROBE, and the score is proportional to the strength of the interaction [31]. PROBE identifies the contacts between residues in a protein by rolling a small virtual probe (0.25 Å) around the van der Waals surface of each atom; an interaction is detected if the probe touches another noncovalently bonded atom. The contact scores are evaluated per dot, and then summed for each atom pair. The combined I_s is the weighted sum of the nonoverlapping van der Waals contacts, the volume of hydrogen bonds, and atomic overlaps (clashes). In contrast to the RINs that are based on the spatial atomic distance cut-off between connecting residues, RINERATOR is capable of generating a more realistic RIN by sampling the atomic packing of each atom by using the small-probe contact dot surface after the inclusion of hydrogen atoms [30].

Construction of subnetworks based on the strength of the noncovalent I_s

We constructed different subnetworks, based on the strength of I_s as described above, between all pairs of residues in which any pair of residues is connected by an edge, if their I_s is higher than a threshold value (I_{smin}). Then, RINs of all three structures (V1L, RBSX, and V1A) were constructed at different I_{smin} values, and their network topology and various network parameters were evaluated and compared. As I_{smin} increases, the number of edges in the RINs decreases, owing to the presence of fewer edges with high I_s values. Once edges with an I_s below the cut-off are removed, nodes with no edges incident on them are no longer considered. All of these networks were visualized using CYTOSCAPE [35]. The NETWORK ANALYZER [53] plugin of CYTOSCAPE was used to calculate simple topological parameters, and the RINALYZER [30] plug-in was used for the comparison of RINs of V1L and V1A structures.

Strongly connected components

Furthermore, we calculated strongly connected components by use of the BINOM2.5 [54] module in CYTOSCAPE for networks of more stable and less stable mutants at different I_{smin} values to identify their distinct clusters and cluster-forming nodes. BINOM uses the algorithm of Tarjan to decompose the network into strongly connected components [55]. The giant cluster (defined here as the LSCC) is the largest group of connected nodes (in terms of number of residues) in the network that are connected to each other. The size of the LSCC (number of nodes) in the network depends on the existence of edges retained based on the I_{smin} between the nodes. Hence, the size of the LSCC can be considered as a function of I_{smin}. We then calculated the size of the LSCC by varying the I_{smin}, and plotted the LSCC as a function of I_{smin}. Here, the size of the LSCC is normalized with respect to the total number of residues in the protein.

Acknowledgements

Author contributions

P. Mahanta, V. S. Reddy and S. Ramakumar designed the experiments. P. Mahanta, A. Bhardwaj and K. Kumar performed the experiment. P. Mahanta crystallized and solved the structures. P. Mahanta and S. Ramakumar carried out the computational analysis. All authors wrote and approved the manuscript.

References

1 Eijsink VGH, Gåseidnes S, Borchert TV & van den Burg B (2005) Directed evolution of enzyme stability. *Biomol Eng* **22**, 21–30.

2 Wijma HJ, Floor RJ & Janssen DB (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr Opin Struct Biol* **23**, 588–594.

3 Baker M (2011) Protein engineering: navigating between chance and reason. *Nat Methods* **8**, 623–626.

4 Shoichet K, Baase WA, Kuroki R & Matthews RW (1995) A relationship between protein stability and protein function. *Proc Natl Acad Sci USA* **92**, 452–456.

5 Korkegian A, Black ME, Baker D & Stoddard BL (2005) Computational thermostabilization of an enzyme. *Science* **308**, 857–860.

6 Joo JC, Pohkrel S, Pack SP & Yoo YJ (2010) Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity. *J Biotechnol* **146**, 31–39.

7 Beg QK, Kapoor M, Mahajan L & Hoondal GS (2001) Microbial xylanases and their industrial applications: a review. *Appl Microbiol Biotechnol* **56**, 326–338.

8 Kamal Kumar B, Balakrishnan H & Rele MV (2004) Compatibility of alkaline xylanases from an alkaliphilic Bacillus NCL (87-6-10) with commercial detergents and proteases. *J Ind Microbiol Biotechnol* **31**, 83–87.

9 Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW & Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804–807.

10 Gírio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S & Bogel-Łukasik R (2010) Hemicelluloses for fuel ethanol: a review. *Bioresour Technol* **101**, 4775–4800.

11 Manikandan K, Bhardwaj A, Gupta N, Lokanath NK, Ghosh A & Reddy VS (2006) Crystal structures of native and xylosaccharide-bound alkali thermostable xylanase from an alkalophilic Bacillus sp. NG-27: structural insights into alkalophilicity and implications for adaptation to polyextreme conditions. *Protein Sci* **91**, 1951–1960.

12 Giver L, Gershenson A, Freskgard PO & Arnold FH (1998) Directed evolution of a thermostable esterase. *Proc Natl Acad Sci USA* **95**, 12809–12813.

13 Wierenga RK (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett* **492**, 193–198.

14 Nagano N, Orengo CA & Thornton JM (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* **321**, 741–765.

15 Gerlt JA & Raushel FM (2003) Evolution of function in (β/α)8-barrel enzymes. *Curr Opin Chem Biol* **7**, 252–264.

16 Bhardwaj A, Leelavathi S, Mazumdar-Leighton S, Ghosh A, Ramakumar S & Reddy VS (2008) The critical role of partially exposed N-terminal valine residue in stabilizing GH10 xylanase from Bacillus sp.NG-27 under poly-extreme conditions. *PLoS One* **3**, e3063.

17 Jacob E & Unger R (2006) A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics* **23**, 225–230.

18 Merz A, Knöchel T, Jansonius JN & Kirschner K (1999) The hyperthermostable indoleglycerol phosphate synthase from *Thermotoga maritima* is destabilized by mutational disruption of two solvent-exposed salt bridges. *J Mol Biol* **288**, 753–763.

19 Notomista E, Catanzano F, Graziano G, Di Gaetano S, Barone G & Di Donato A (2001) Contribution of chain termini to the conformational stability and biological activity of onconase. *Biochemistry* **40**, 9097–9103.

20 Liu L, Zhang G, Zhang Z, Wang S & Chen H (2011) Terminal amino acids disturb xylanase thermostability and activity. *J Biol Chem* **286**, 44710–44715.

21 Bhardwaj A, Mahanta P, Ramakumar S, Ghosh A, Leelavathi S & Reddy VS (2012) Emerging role of N- and C-terminal interactions in stabilizing (β/α)8 fold with special emphasis on Family 10 xylanases. *Comput Struct Biotechnol J* **2**, e20120914.

22 Song L, Tsang A & Sylvestre M (2015) Engineering a thermostable fungal GH10 xylanase, importance of N-terminal amino acids. *Biotechnol Bioeng* **112**, 1081–1091.

23 Krishna MMG & Englander SW (2005) The N-terminal to C-terminal motif in protein folding and function. *Proc Natl Acad Sci USA* **102**, 1053–1058.

24 Kabsch W & Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.

25 Sterner R & Höcker B (2005) Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem Rev* **105**, 4038–4055.

26 Mamo G, Thunnissen M, Hatti-Kaul R & Mattiasson B (2009) An alkaline active xylanase: insights into mechanisms of high pH catalytic adaptation. *Biochimie* **91**, 1187–1196.

27 Collins T, Gerday C & Feller G (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev* **29**, 3–23.

28 Brinda KV & Vishveshwara S (2005) A network representation of protein structures: implications for protein stability. *Biophys J* **89**, 4159–4170.

29 Di Paola L, De Ruvo M, Paci P, Santoni D & Giuliani A (2013) Protein contact networks: an emerging paradigm in chemistry. *Chem Rev* **113**, 1598–1613.

30 Doncheva NT, Klein K, Domingues FS & Albrecht M (2011) Analyzing and visualizing residue networks of protein structures. *Trends Biochem Sci* **36**, 179–182.

31 Doncheva NT, Assenov Y, Domingues FS & Albrecht M (2012) Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc* **7**, 670–685.

32 Kinjo AR & Nishikawa K (2005) Recoverable one-dimensional encoding of three-dimensional protein structures. *Bioinformatics* **21**, 2167–2170.

33 Greene LH & Higman VA (2003) Uncovering network systems within protein structures. *J Mol Biol* **334**, 781–791.

34 Sengupta D & Kundu S (2012) Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. *BMC Bioinformatics* **13**, 142.

35 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504.

36 Zhang X, Perica T & Teichmann SA (2013) Evolution of protein structures and interactions from the perspective of residue contact networks. *Curr Opin Struct Biol* **23**, 954–963.

37 Damodaran A (1991) Tragedy of the commons and comedy of common property resources. *Econ Polit Wkly* **26**, 2213–2215.

38 Zhou R, Eleftheriou M, Royyuru AK & Berne BJ (2007) Destruction of long-range interactions by a single mutation in lysozyme. *Proc Natl Acad Sci USA* **104**, 5824–5829.

39 Rader AJ, Yennamalli RM, Harter AK & Sen TZ (2012) A rigid network of long-range contacts increases thermostability in a mutant endoglucanase. *J Biomol Struct Dyn* **30**, 628–637.

40 Kamondi S, Szilágyi A, Barna L & Závodszky P (2008) Engineering the thermostability of a TIM-barrel enzyme by rational family shuffling. *Biochem Biophys Res Commun* **374**, 725–730.

41 Bhardwaj A, Leelavathi S, Mazumdar-Leighton S, Ghosh A, Ramakumar S & Reddy VS (2010) The critical role of N- and C-terminal contact in protein stability and folding of a family 10 xylanase under extreme conditions. *PLoS One* **5**, e11347.

42 Deng Z, Huang W, Bakkalbasi E, Brown NG, Adamski CJ, Rice K, Muzny D, Gibbs RA & Palzkill T (2012) Deep sequencing of systematic combinatorial libraries reveals β-lactamase sequence constraints at high resolution. *J Mol Biol* **424**, 150–167.

43 Delbrück H, Mueller U, Perl D, Schmid FX & Heinemann U (2001) Crystal structures of mutant forms of the *Bacillus caldolyticus* cold shock protein differing in thermal stability. *J Mol Biol* **313**, 359–369.

44 Perl D & Schmid FX (2001) Electrostatic stabilization of a thermophilic cold shock protein. *J Mol Biol* **313**, 343–357.

45 Buckle AM, Henrick K & Fersht AR (1993) Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J Mol Biol* **234**, 847–860.

46 Takano K, Scholtz JM, Sacchettini JC & Pace CN (2003) The contribution of polar group burial to protein stability is strongly context-dependent. *J Biol Chem* **278**, 31790–31795.

47 Krobath H, Rey A & Faísca PFN (2015) How determinant is N-terminal to C-terminal coupling for protein folding? *Phys Chem Chem Phys* **17**, 3512–3524.

48 Szilágyi A & Závodszky P (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493–504.

49 Berezovsky IN & Shakhnovich EI (2005) Physics and evolution of thermophilic adaptation. *Proc Natl Acad Sci USA* **102**, 12742–12747.

50 Orengo CA, Jones DT & Thornton JM (1994) Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.

51 Word JM, Lovell SC, Richardson JS & Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735–1747.

52 Word JM, Lovell SC, Labean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS & Richardson DC (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* **285**, 1711–1733.

53 Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T & Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284.

54 Zinovyev A, Viara E, Calzone L & Barillot E (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics* **24**, 876–877.

55 Tarjan RE (1983) Depth-first serach and linear graph algorithms. *SIAM J Comput* **1**, 146–460.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Doc. S1. Supporting information.

Fig. S1. RWC$_\beta$CO of RBSX and different mutants.

Fig. S2. Unique N-terminal to C-terminal contacts for less stable (V1A) and more stable (V1L) mutants.

Fig. S3. Structural superpositions of TIMs and glyceraldehyde-3-phosphate dehydrogenases (GAPDHs) from different organisms.

Table S1. Van der Waals contacts of mutated residues.

Table S2. Unique N-terminal to C-terminal atom–atom contacts.

Table S3. Unique residues in the cluster of the LSCC.

Table S4. Comparison of N-terminal to C-terminal contacts of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) structures from different organisms.