The FEBS Journal

FEBS PRESS science publishing by scientists

Loop size optimization induces a strong thermal stabilization of the thioredoxin fold

Alessia Ruggiero[1], Giovanni Smaldone[2], Luciana Esposito[1], Nicole Balasco[1] and Luigi Vitagliano[1]

1 Institute of Biostructures and Bioimaging, C.N.R., Naples, Italy
2 IRCCS SDN, Naples, Italy

Keywords
loop size optimization; protein engineering; protein thermal stability; protocol of protein stabilization; thioredoxin chimeric proteins

Correspondence
L. Vitagliano and A. Ruggiero, Institute of Biostructures and Bioimaging, C.N.R., Via Mezzocannone 16, Naples, Italy
Tel: +0039 0812534506
E-mails: luigi.vitagliano@unina.it (LV), alessia.ruggiero@unina.it (AR)

(Received 3 September 2018, revised 21 December 2018, accepted 22 January 2019)

doi:10.1111/febs.14767

The definition of the structural basis of protein thermostability represents a major topic in structural biology and protein chemistry. We have recently observed that proteins isolated from thermophilic organisms show a better adherence to the fundamental rules of protein topology previously unveiled by Baker and coworkers (Koga *et al.* Nature. 2012; **491**: 222–227). Here, we explored the possibility that *ad hoc* modifications of a natural protein following these rules could represent an efficient tool to stabilize its structure. Hence, we here designed and characterized novel variants of *Escherichia coli* thioredoxin (*Ec*Trx) using a repertoire of biophysical/structural techniques. Trx chimeric variants were prepared by replacing the loop of *Ec*Trx with the corresponding ones present in the Trxs isolated from *Sulfolobus solfataricus* and *Sulfolobus tokodaii* that show a better adherence to the topological rules. Interestingly, although the loop sequences of these proteins did not display any significant similarity, their insertion in *Ec*Trx induced a remarkable stabilization of the protein (≥10 °C). The crystallographic structure of one of these variants corroborates the hypothesis that the optimization of the loop size is the driving force of the observed stabilization. The remarkable stabilization of the two novel chimeric Trxs, generated by applying the topological rules, represents the proof of concept that these rules may be used to stabilize natural proteins through the *ad hoc* optimization of the loop size. Based on the present results, we propose a novel protocol of protein stabilization that can be potentially applied to other proteins.

Introduction

The most important unsolved problem in protein science is the deciphering of the folding code that links protein three-dimensional structures to their sequences [1–3]. Although significant advances in the field have been made in the last decades, successful *ab initio* predictions of protein structures are sporadic, although important [4–9]. It should also be considered that most of these successful approaches rely more on the analysis of the structural preferences of protein fragments deduced from experimental databases than on comprehensive and accurate evaluations of the basic physicochemical principles that govern folding and fold [10].

Even more difficult is the identification of the principles that confer protein structures elevated stabilities

Abbreviations
CD, circular dichroism; *Ec*Trx, thioredoxin isolated from *Escherichia coli*; PDB, protein data bank; *Ss_Ec*Trx, chimera of EcTrx whose residues 14–22 have been replaced with residues 40–47 of TrxA1 from *S. solfataricus*; *Ss*TrxA1, thioredoxin A1 isolated from *Sulfolobus solfataricus*; *St_Ec*Trx, chimera of EcTrx whose residues 14–22 have been replaced with residues 47–54 of TrxA1 from *S. tokodai*; *St*Trx, thioredoxin isolated from *Sulfolobus tokodaii*; Trx, thioredoxin.

when exposed to unusual external conditions (e.g., very high/low temperature or chemical denaturants) [11–17]. Although the development of proteins endowed with enhanced stabilities would have a tremendous impact in chemistry, biology, and medicine, their rational design is still an elusive goal [18–21]. Apparently, the extremely high thermal stability of proteins isolated from thermophilic organisms, whose growth temperatures exceed 80 °C, originates from a rich mixture of distinct factors. Indeed, several mechanisms underlying protein thermal stability have been proposed including higher packing efficiency through van der Waals interactions, networks of ion pairs and/or hydrogen bonds, reduction in conformational strain and others [13,22–24]. Among these, loop shortening is considered a widespread mode used by proteins from thermophilic organisms to achieve stabilization [25]. Since protein stabilization relies on the fine balance of forces that favor either the folded or the unfolded state, it is commonly believed that current difficulties in the field arise from our limited knowledge of the structural features of the unfolded state. In a seminal recent work, Baker and colleagues have shown that the lengths of loop connecting secondary structure elements is, in stable folds, optimized to follow well-defined rules [26]. The rationale behind these principles, also denoted as fundamental rules, is that specific loop lengths may favor the folded state by reducing the number of states and, hence, the resulting entropy of the unfolded state. Notably, these authors were also able to generate hyperstable domain scaffolds that were specifically designed to follow these rules. On the basis of these findings, we hypothesized that natural proteins endowed with high thermal stabilities have somehow evolved to adhere to these rules [27]. By comparing proteins isolated from organisms living in different environmental conditions, we indeed showed that those extracted from thermophilic organisms better follow these rules when compared to those isolated from mesophilic organisms [27]. These findings prompted us to suppose that the length optimization of loops connecting secondary structure elements following the rules of the Baker group could be a valuable way to stabilize protein structures. We here evaluated the effect of this loop length optimization strategy using the thioredoxin isolated from *Escherichia coli* (*Ec*Trx) as model system. *Ec*Trx is a prototypal system for the analysis of structure/function/stability relationships [28–30]. This protein has also been frequently engineered to be adapted to perform specific functions (scaffold for epitope presentation, fusion tag, etc.) [29,31–34]. Not surprisingly, these *Ec*Trx modifications often induce a certain level of structural perturbation that results in significant destabilizations. In this scenario, Trx variants with enhanced stability represent better supports for *ad hoc* modifications [35,36]. By adopting the length loop optimization strategy, we here designed and characterized two novel *Ec*Trx variants endowed with a remarkable thermostability. On the basis of these results, we also proposed a novel protocol of protein stabilization that can be applied to other protein systems.

Results and Discussion

Protein design

Previous comparative analyses of the loop size of Trx structures isolated from thermophilic (*Sulfolobus solfataricus, Sulfolobus tokodaii, Thermus thermophilus*) or mesophilic (*Escherichia coli, Spinach chloroplast*) sources have shown that the loop connecting helix 1 (α1) to strand 2 (β2) in the first αℓβ motif of thioredoxin proteins undergoes a shortening from six/seven (mesophilic organisms) to two residues (thermophilic organisms) [27,37,38]. According to the Baker's fundamental rules [26], αℓβ-ℓ6 and αℓβ-ℓ7 states (see Materials and Methods section for the notation) do not show any preference between the two possible P (parallel) or A (antiparallel) orientations of the two connected elements (helix and strand). The αℓβ-ℓ2 state displays a strong preference for the P-orientation that is the one observed in the Trx fold. In this framework, the shortening of the loop size may favor the formation of the native P-oriented α1-β2 motif in Trx thus stabilizing the proteins extracted from the thermophilic sources. On the basis of these considerations, we designed chimeric variants of *Ec*Trx by replacing the loop of its α1-β2 motif (residues 14–22) with the corresponding αℓβ-ℓ2 loops of two thioredoxin proteins isolated from thermophilic sources. The regions to be swapped were identified upon local sequence alignments that highlighted regions with high sequence conservation (secondary structure elements) and region with elevated sequence variability (loop) (Fig. 1). The chimeric proteins were generated by replacing the α1-β2 loop region (residues 14–22) of *E. coli* Trx with corresponding ones from *S. solfataricus* TrxA1 (*Ss*TrxA1) and from *S. tokodai* Trx (*St*Trx). In particular, the loop regions inserted in the *E. coli* scaffold to generate the chimeras were residues 40–47 from *Ss*TrxA1 (*Ss_Ec*Trx chimera) and residues 47–54 from *St*Trx (*St_Ec*Trx chimera). It is important to note that, in the two Trx proteins from thermophiles, the α1-β2 loops have completely different sequences as only one residue out of eight is conserved (Figs 1B,C and 2).

These analyses were also expanded by considering, as a sort of negative control, a loop isolated from a thermophilic source with a size identical to that of *E. coli* Trx. To this aim, we considered the α3-β4 loop which has the same length in EcTrx and SsTrxA1 (Fig. 1). This loop is also too long (seven residues—αℓβ-ℓ7) to have an impact on the fold stability according to the topology rules (Fig. 1D). A new chimeric protein, *Ss2_Ec*Trx, was generated by replacing the α3-β4 loop region (residues 70–76) of *E. coli* Trx with the corresponding one of *S. solfataricus* (TrxA1 residues 94–100).

Finally, we also evaluated the impact of inserting in *E. coli* Trx scaffold the sequence of the α1-β2 loop extracted from the Trx protein of the mesophilic organism *S. aureus* (*Sa*Trx). As shown in Fig. 1E, the α1-β2 loop of *Sa*Trx presents a reduced size (three residues) compared to *E. coli* Trx [39]. Although the size of this loop (αℓβ-ℓ3) in *Sa*Trx does not favor the P-orientation required by the Trx fold [26], we inserted it in *Ec*Trx since this may lead to a shorter α1-β2 loop. In this case, the chimeric protein *Sa_Ec*Trx was generated by replacing the α1-β2 loop region (residues 14–22) of *E. coli* with the corresponding one of *S. aureus* (residues 13–19).

Protein production and biophysical characterization

The chimeric proteins (*Ss_Ec*Trx, *St_Ec*Trx, *Ss2_Ec*Trx, and *Sa_Ec*Trx) were expressed and purified in soluble and homogeneous forms that were amenable to both solution and solid state studies (see Materials and Methods for details). Initial experiments in solution were performed on the oxidized forms of the proteins by carrying out the experiments under air atmosphere. As shown in Fig. 3A, the Far-UV CD spectra of all chimeric proteins closely resemble the one exhibited by *Ec*Trx. This suggests that the insertion of different sequences of α1-β2 or α3-β4 loops in *Ec*Trx does not perturb the overall fold of these variants.

In order to gain insights into the structural stability of the chimeric proteins against thermal denaturation, we evaluated the variation in the CD signal upon temperature increase (Fig. 3B). In line with literature data [32], the oxidized form of *Ec*Trx presents a melting temperature (Tm) of 82 °C. The inspection of the denaturation curves clearly indicates that *Ss_Ec*Trx and *St_Ec*Trx chimeras are characterized by a significantly increased thermostability compared to the parent protein as both of them exhibit Tm values higher than 90 °C (Table 1). The thermal denaturation process of these chimeric variants is reversible as indicated by the recovery of CD signal of the cooled sample that was previously denatured (Fig. 4A–D).

Similar results are obtained when the reduced forms of these proteins are characterized. In agreement with literature data [40], the reduced state of *Ec*Trx presents a significantly lower stability (Tm of 70 °C) if compared to its oxidized form (Table 1). Again, the replacement of the α1-β2 loop has a remarkable impact of the stability of the protein also in its reduced state. Indeed, the Tm of *Ss_Ec*Trx and *St_Ec*Trx chimeras is larger than 80 °C (Fig. 3C and Table 1).

As for the oxidized states of the chimeras, these unfolding processes are also reversible as highlighted by the comparison of the spectra of the starting and of the cooled samples. These data indicate that both *Ss_Ec*Trx and *St_Ec*Trx, designed by simply following the fundamental rules, are remarkably more stable than the wild-type protein despite their completely different sequence of the α1-β2 loop.

To check whether the observed effect was not merely due to the insertion of a loop extracted from a thermophilic organism within the scaffold of a mesophilic protein, we also characterized the variant *Ss2_Ec*Trx in which the loop α3-β4 of *Ec*Trx was replaced with the corresponding region of SsTrxA1 (see above and Fig. 1D). As shown in Fig. 3B,C, no stabilization was observed for this variant compared to the parent protein. This indicates that the simply insertion of a fragment from a thermophilic source within the *Ec*Trx scaffold does not *per se* warrant stabilization.

Finally, the replacement of the *Ec*Trx α1-β2 loop with the corresponding one from a mesophilic organism *S. aureus* (*Sa_Ec*Trx) only produced marginal stabilizations (Fig. 3B,C, Table 1).

Crystallographic studies

To verify the full adherence to the fundamental rules of the α1-β2 loop of these variants, we undertook structural studies aimed at elucidating the conformation of this region. In this framework, we set up crystallization trials on all the chimeric EcTrx forms designed. Unfortunately, crystals suitable for crystallographic studies were obtained only for *Ss_Ec*Trx. Indeed, for this variant, the X-ray crystallographic structure was determined at 2.99 Å resolution (Fig. 5). Despite the limited resolution, the refined model presents rather good stereochemistry quality (see Methods for details). Moreover, the quality of the electron density of the α1-β2 region allows a reliable definition of the conformation adopted by the residues embodied in

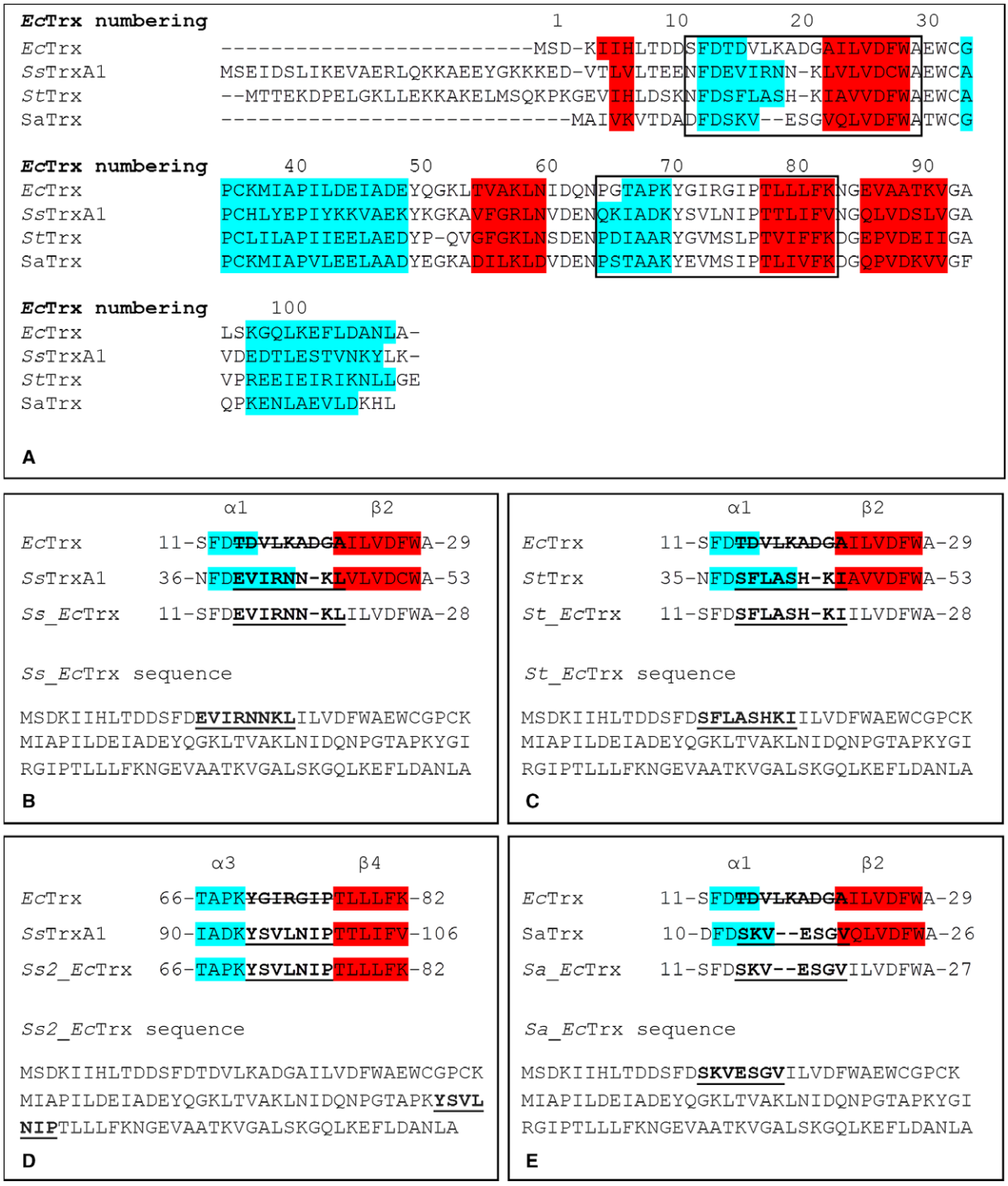

Fig. 1. Sequence alignment of Trx proteins isolated from different organisms (*E. coli, S. solfataricus, S. tokodaii, and S. aureus*) (A). The boxes identify the α1-β2 and α3-β4 motifs. Sequence of the region carrying the α1-β2 or α3-β4 motifs in the chimeric variants *Ss_Ec*Trx (B), *St_Ec*Trx (C), *Ss2_Ec*Trx (D), and *Sa_Ec*Trx (E). The α-helices and β-strands of the crystal structures (*Ec*Trx: 2TRX PDB code, *Ss*TrxA1: 3TCO PDB code, *St*Trx: 2E0Q PDB code, and *Sa*Trx: 2O7K PDB code) are highlighted in cyan and red, respectively. Secondary structure assignment has been performed using DSSP.

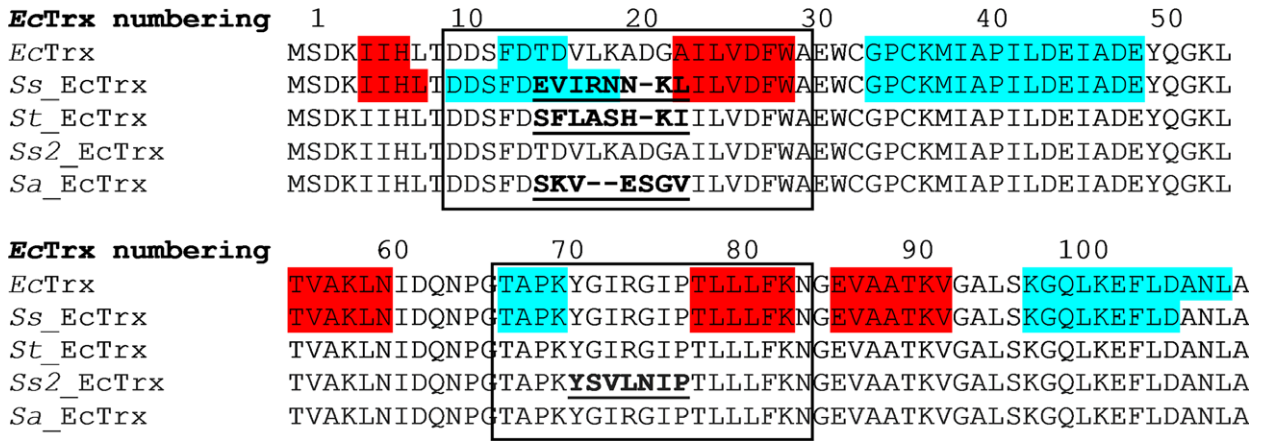

Fig. 2. Full-length alignment of the entire sequences of *Ec*Trx, *Ss_Ec*Trx, *St_Ec*Trx, *Ss2_Ec*Trx, and *Sa_Ec*Trx. The boxes identify the α1-β2 and α3-β4 regions. The α-helices and β-strands of *Ec*Trx (2TRX PDB code) and *Ss_Ec*Trx (6H1Y PDB code), assigned by DSSP, are highlighted in cyan and red, respectively.

this region. Indeed, as shown in Fig. 6, with the exception of the very terminal portion of Lys20 side chain, the electron density is generally well defined. The assignment of the secondary structure performed using the program DSSP indicates that the α1 helix ends with residue Asn18 whereas the β2 starts with residue Leu21 (Fig. 1B). These observations clearly indicate that the α1-β2 loop in *Ss_Ec*Trx is made of only two residues (Asn19-Lys20). Therefore, compared to the wild-type protein, the loop size in this chimera reduces from six to two residues. This reduction is the result of the deletion of one residue in the sequence and of a local rearrangement of the protein chain that produces an enlargement of the α1 helix. Indeed, as shown in Fig. 7, the local structure of *Ss_Ec*Trx is more similar to *Ss*TrxA1 than to *Ec*Trx, although the rest of the sequence is the *Ec*Trx one. As a result of the rearrangement of the local structure, the α1-β2 loop of this chimera is of the type $\alpha\ell\beta$-$\ell 2$ which strongly favors the P association of the helix and the strand that is one adopted in the Trx fold.

It is important to note that, in principle, this chimera could be stabilized by strong favorable interactions made by the residues of this loop with the rest of the protein. The inspection of the local structural context of the side chains of residues in the loop clearly indicates that they do not make major contacts with the rest of the protein main body. The only exception is a rather weak H-bond between the $N^{\delta 2}$ atom of Asn19 side chain with the main chain O atom of Val15 (distances 3.5 Å and 2.6 Å in the two mates present in the asymmetric unit). In this context, it is worth mentioning that stabilizing effects are also produced by the replacement of *Ec*Trx loop with the *St*Trx one, which instead presents a His residue in position 19 (Fig. 2). In conclusion, being the loop sequence of *Ss*TrxA1 and *St*Trx very different, it is unlikely that both chimeras are stabilized by interactions made by the loop region with the rest of the protein. These considerations indicate that topological aspects have an important impact on Trx stability. Indeed, the remarkable stabilization of two different variants of *Ec*Trx, generated by applying the concepts underlying the fundamental rules in a single mutational event, represents the proof of concept that these rules may be used to stabilize natural proteins through the *ad hoc* optimization of the loop size. The marginal stabilization obtained in the case of *Sa_Ec*Trx cannot be fully interpreted in the absence of a 3D structure for this variant. Since variations of even a single residue within a loop may change its impact of the stabilization of a specific topology, it is important to evaluate its size in the structured fold. Indeed, the lack of information on the size of the loop in this structure precludes the possibility to verify whether this region follows the topology rules.

An innovative and general protein stabilization protocol through loop size optimization

Present findings suggest that loop optimization strategies used to develop protein variants that (fully) adhere to the fundamental rules of protein topology represent an innovative and powerful tool to enhance the thermal stability of natural and/or artificial proteins. On the basis of the present results, we propose a novel protocol of protein stabilization that can be

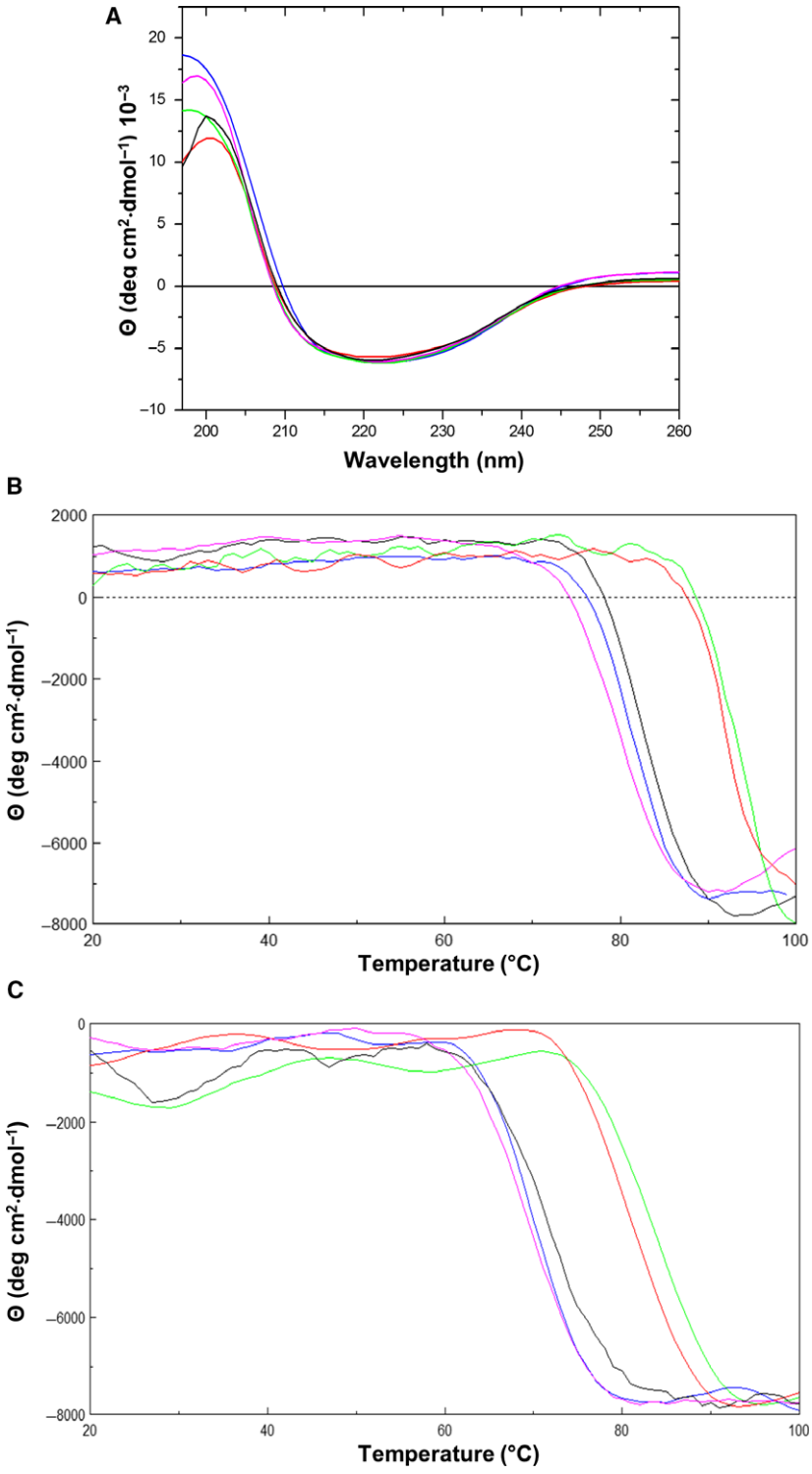

Fig. 3. Far-UV CD spectra of the oxidized forms of *Ec*Trx (blue), *Ss_Ec*Trx (green), *St_Ec*Trx (red), *Ss2_Ec*Trx (magenta), and *Sa_Ec*Trx (black) (A). Thermal denaturation curves of the wild-type protein *Ec*Trx and the chimeric forms *Ss_Ec*Trx, *St_Ec*Trx, *Ss2_Ec*Trx, and *Sa_Ec*Trx in the oxidized (B) and reduced (C) forms. The change in ellipticity [θ] at 208 nm when the temperature increases from 20 °C to 100 °C was monitored.

applied to a large variety of proteins. As detailed below the protocol can be widely applied, although the chances of success are likely increased by the availability of structural data on the investigated protein and/or the availability of sequence/structural data on thermostable homologs. As shown in Fig. 8, this protocol is organized in two scenarios, each comprising three different cases, as follows.

Table 1. Melting temperatures of EcTrx and the chimeric variants in the oxidized and reduced states obtained from CD data.

Protein	T_m (°C) (oxidized state)	T_m (°C) (reduced state)
*Ec*Trx	82	70
*Ss_Ec*Trx	93	84
*St_Ec*Trx	92	81
*Ss2_Ec*Trx	80	70
*Sa_Ec*Trx	84	72

Scenario 1—The three-dimensional structure of the investigated protein is known

Case (a): The three-dimensional structure of a thermostable homolog is known

This represents the most promising case. The first step is a comparative analysis of the loop sizes of both proteins. This comparison aims at unraveling differences in the loop adherence to the fundamental rules. If the thermostable counterpart exhibits specific loops that better follow the rules, chimeric variants could be generated by replacing the sequences of these loops in the original sequence of the investigated protein.

Case (b): Only the sequence of a thermostable homolog is known

If only the structure of the investigated protein is known, sequence alignments and secondary structure prediction will be applied to locate the loops in the thermostable counterpart. Once loop regions have been defined in both proteins, the stabilization of the investigated protein will be pursued following the approach described in case A (identification of insertions/deletion and generation of chimeric variants). A similar reversed approach can be used if only the structure of the thermostable protein is known (Scenario 2: *case a*).

Case (c): Thermostable homologs are unknown

The loop sizes of the protein will be analyzed to evaluate their adherence to the rules. Variants of the protein will be designed through the optimization of the loop sizes for a better adherence to the rules.

Scenario 2—The three-dimensional structure of the investigated protein is unknown

Case (a): The three-dimensional structure of a thermostable homolog is known

If the structure of a thermostable homolog of the investigated protein is known, sequence alignments and secondary structure prediction will be applied to locate the loops in the investigated counterpart. Once loop regions have been defined in both proteins, the stabilization of the investigated protein will be pursued following the approach described in scenario 1—case

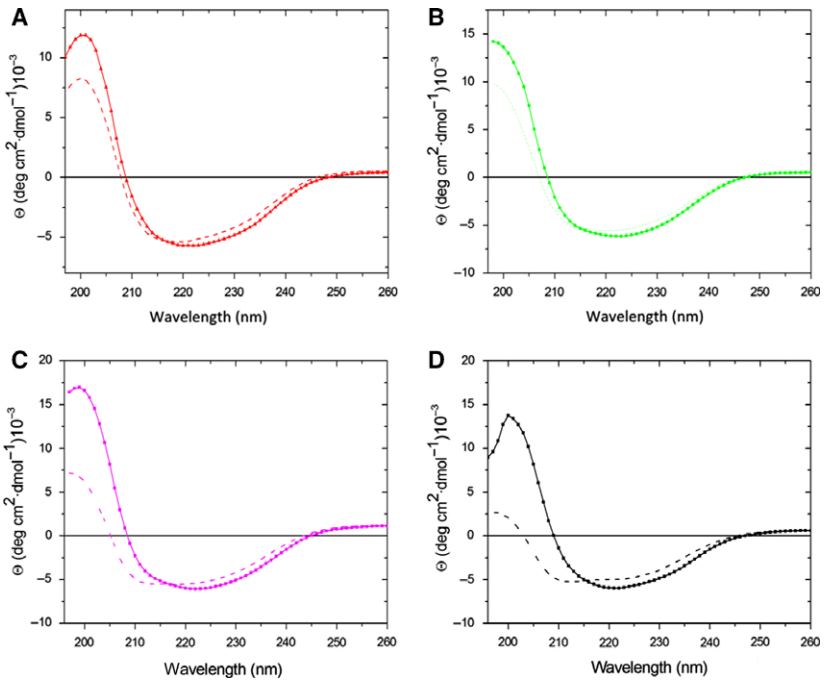

Fig. 4. Far-UV CD spectra of the oxidized forms of *Ss_Ec*Trx (A), *St_Ec*Trx (B), *Ss2_Ec*Trx (C), and *Sa_Ec*Trx (D) at the starting temperature of 20 °C (straight line) and of the cooled samples after melting (dashed line).

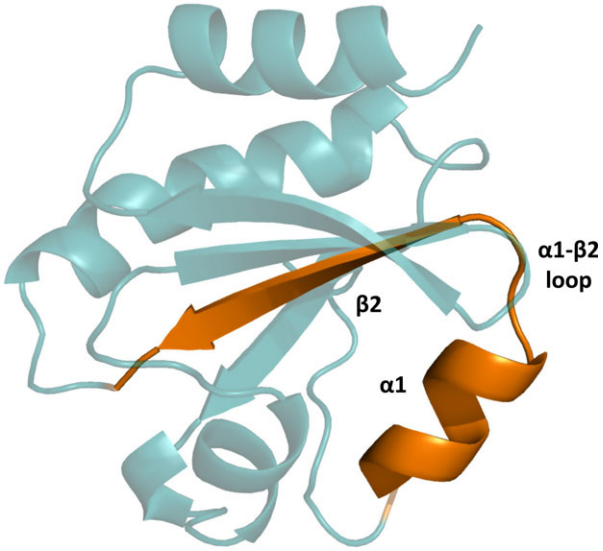

Fig. 5. Cartoon representation of the chimeric variant *Ss_Ec*Trx solved by X-ray crystallography at 2.99 Å resolution. The α1 helix and the β2 strand are shown in orange.

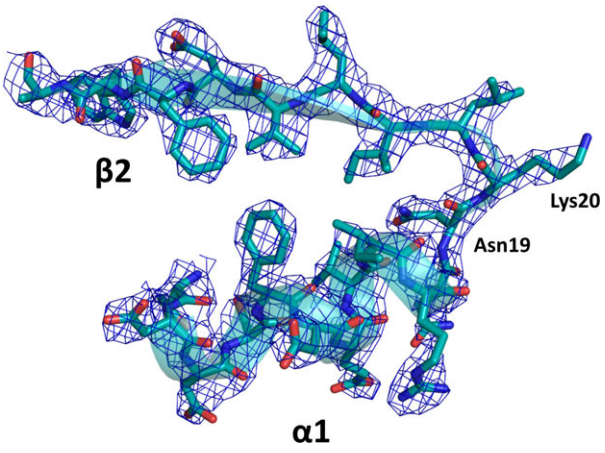

Fig. 6. |2Fo – Fc| electron density maps (contoured at 1.5 σ) of the residues of the α1-β2 region (residues 8–28) in *Ss_Ec*Trx.

(*a*) (identification of insertions/deletion and generation of chimeric variants).

Case (b): Only the sequence of a thermostable homolog is known

In this case, information on the location of the loops in the two sequences will be obtained performing secondary structure predictions. The occurrence of insertions/deletions in loop regions will then be evaluated. Since the rules suggest that the impact of loop size on the overall topology is larger for shorter loops, the attention will be focused on loops that are shorter in the thermostable protein. If any, they will be inserted in the investigated protein by replacing the original ones.

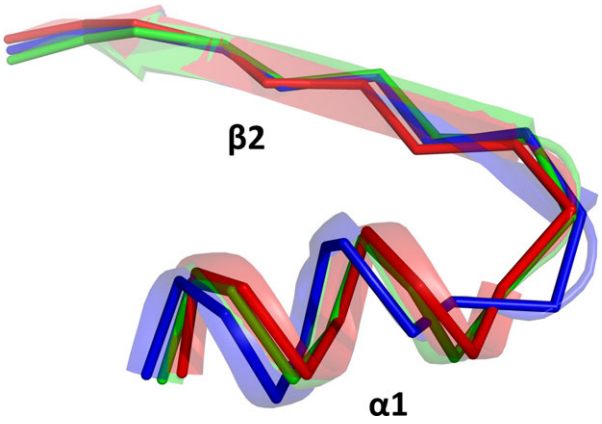

Fig. 7. Structural alignment of the α1-β2 motif of *Ec*Trx (blue—2TRX PDB code), *Ss*TrxA1 (red—3TCO PDB code), and *Ss_Ec*Trx (green—6H1Y PDB code). The trace representation is used.

Case (c): Thermostable homologs are unknown

This is evidently the worst case. Information on the location of the loops in the sequence can be obtained through secondary structure predictions. Then, loop regions may be tentatively shortened to increase the impact of the loop size on the association modes of the linked secondary structure elements. If this loop shortening favors the (unknown) topology of the fold, then this mutation is likely to increase the stability of the variant.

Concluding remarks

Protein thermal stabilization represents a valuable subject of both academic and industrial research. Thermally stable proteins have a wide range of practical applications as they allow more efficient catalysis in industrial processes and the improvement of the efficacy and shelf-life of protein-based pharmaceuticals [18–21,41]. High thermal stability makes both protein storage and handling easier and allows some high-temperature reactions. A wide number of experimental and theoretical approaches have been developed aimed at engineering proteins endowed with a remarkable thermal stability [15–21,35]. However, a full comprehension of the basic principles that govern protein thermal stability represents one of the main goals of structural biology. In this framework, our analysis provides the first experimental support to the innovative concept of protein thermal stabilization based on the optimization of the loop size to better adhere to the fundamental rules. On the basis of the present

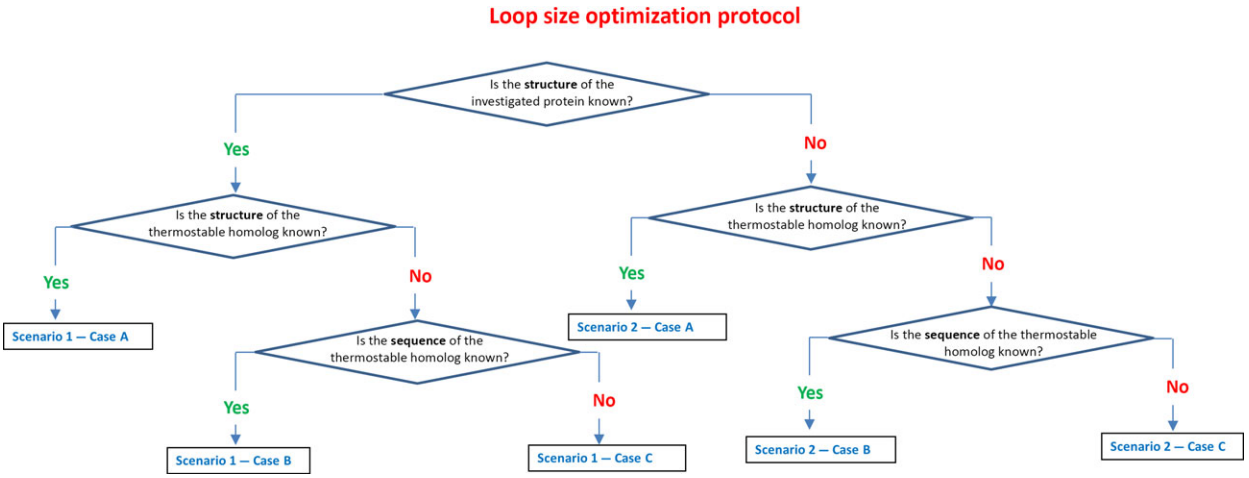

Fig. 8. Scheme of the loop size optimization protocol. Cases are described in the text.

results, we also proposed a novel protocol of protein stabilization that can be applied in different scenarios as function of the level of structural information available. Although this protocol requires further validation, we believe that it may represent an effective option in designing thermostable proteins. Finally, present studies led to the generation of novel thermostable variants of *E. coli* Trx, a protein frequently used in biotechnological applications as scaffold or tag [28–30], that could be further exploited in the future in both basic and applied studies.

Materials and methods

Notation

Taking into account the notation we previously used in Balasco *et al.* [27], loops linking an α to a β element are here denoted as αℓβ-ℓx where x corresponds to the number of the residues that constitute the loop.

Reagents and materials

Restriction enzymes were purchased from New England BioLabs, Inc. (Beverly, MA, USA). Oligonucleotides used for PCR amplification and sequencing were synthesized by Eurofins Genomics. DNA Polymerase, T4 Ligase, *E. coli* strain DH5α competent cells, *E. coli* BL21(DE3) competent cells were purchased from Invitrogen (Invitrogen, Carlsbad, CA, USA).

Cloning, expression, and purification of Ss_EcTrx, St_EcTrx, Ss2_EcTrx, and Sa_EcTrx

The gene encoding the 109 amino acids of *E. coli* thioredoxin (UniProt entry: P0AA25, *Ec*Trx) was amplified by PCR and cloned into pETM-13 expression vector (EMBL) using NcoI and XhoI (New England BioLabs, Inc) restriction sites. Chimeric genes were purchased from GeneWiz (Sigma-Aldrich S.r.l., Milan, Italy) and then subcloned into pETM-13. All protein constructs were endowed with C-terminal poly-Histag that was not cleavable by proteases. *Escherichia coli* BL21 (DE3) strain was used to perform protein expression. Briefly, an overnight culture was used to inoculate 1 l of LB medium containing 50 $\mu g \cdot L^{-1}$ of kanamycin and protein induction was performed by the addition of 0.8 mM IPTG at 298 K. After approximately 16 h, the cells were harvested and the protein was isolated by sonicating cell pellets resuspended in 30 mL of lysis buffer (50 mM Tri-HCl, 200 mM NaCl, pH 7.8) in the presence of a protease-inhibitor cocktail (Roche Diagnostics S.p.A., Monza, Italy). The crude cell extract was cleared by centrifugation at 39 000 ***g*** and the supernatant was loaded onto a 5 mL Ni–NTA column connected to an ÄKTA FPLC system (GE Healthcare Italia, Milan, Italy) equilibrated with binding buffer (50 mM Tris–HCl, 200 mM NaCl, 10 mM imidazole pH 7.8). After washing with 10 volumes of binding buffer, a gradient of imidazole (10–500 mM) was applied to elute the protein. A final purification step was carried out using a Superdex 75 16/60 (GE Healthcare) column (50 mM Tris–HCl, 200 mM NaCl, pH 7.8). The homogeneity of the proteins was confirmed by SDS/PAGE. Freshly concentrated proteins were used for crystallization experiments. The reduced form of *Ec*Trx and the chimeric variants used for biophysical characterization was obtained by using 1.5 mM dithiothreitol (DTT).

Circular dichroism spectroscopy

The correct folding of proteins was assessed by Far-UV CD spectroscopy. The spectra were collected on a Jasco J-715 spectropolarimeter equipped with a Peltier temperature control system (Model PTC-423-S) from 195 to 250 nm at 20 °C using a quartz cell of 0.1-cm path length.

Recordings were carried out at a temperature of 20 °C under constant N_2 flow using the following parameters: scanning speed of 20 nm·min^{-1}, band width of 2 nm, and response time of 4 s. The spectra were recorded at protein concentration 0.3 g·L^{-1}, at pH 7.3 in 20 mM phosphate buffer and they were signal averaged over three scans. The baseline was corrected by subtracting the complete buffer spectrum. The final spectra were expressed as molar ellipticity [θ] (deg cm^2·dmol^{-1}) per residue.

The temperature of the midpoint of the transition (melting Temperature, Tm) was determined by monitoring the change in ellipticity at 208 nm when the temperature increases from 20 °C to 100 °C. This choice was motivated by a comparative analysis of the CD spectra of *Ec*Trx registered at room and high temperatures. As shown in Fig. 9, the difference in the CD signal increases going toward low wavelengths. The reversibility of the transition was then checked by lowering the temperature to 20 °C.

Protein crystallization and crystallographic data collection

Crystallization trials were performed at 293 K using the hanging-drop vapor-diffusion methods. A preliminary screening of the crystallization conditions was carried out by using commercially available sparse-matrix kits (Crystal Screen kits I/II and Index by Hampton Research) [42]. Optimization of the crystallization conditions was performed by fine-tuning the protein and precipitant concentrations. The initial screening of commercially available solutions provided promising crystallization conditions for *Ss_Ec*Trx in the presence of ammonium sulfate as precipitant. The size and the morphology of these small crystals were improved by modulating the crystallization conditions. Crystals that could be tested for diffraction experiments were obtained with a protein concentration of 20–25 mg·mL^{-1}. The composition of the reservoir solution was 17% (w/v) PEG10000, 0.1 M ammonium acetate and 0.1 M BIS-TRIS buffer, pH 5.5. Diffraction data were collected in-house using a Rigaku Micromax 007 HF generator equipped with a Saturn944 CCD detector at 100 K. The data were collected by flash-cooling in the supercooled N_2 gas produced by an Oxford Cryosystem Cryostream after the addition of 20% glycerol to the harvesting solution. The crystals belonged to the P31 space group and diffracted at 2.99Å. The dataset was scaled and merged using the HKL2000 program package (HKL Research, Inc., Charlottesville, VA, USA). Statistics of data collection are reported in Table 2. The crystal structure of *Ss_Ec*Trx was solved by molecular replacement using the program PHASER (Phaser crystallographic software, University of Cambridge, UK) and the structure of thioredoxin from *E. coli* (PDB ID: 2TRX) deprived from α1-β2 loop residues as starting model.

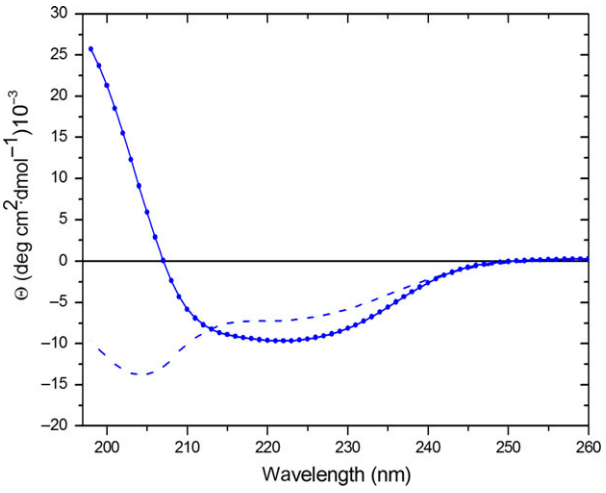

Fig. 9. Far-UV CD spectra of the oxidized form of *Ec*Trx at 20 °C (straight line) and at 100 °C (dashed line).

Structure refinement and validation

Crystallographic refinement of the structure of *Ss_Ec*Trx was carried out against 88% of the measured data using the CCP4I program suite (CCP4, Daresbury Laboratory, Warrington, UK). The remaining 12% of the observed

Table 2. Data collection statistics

	*Ss_Ec*Trx
X-ray device	Rigaku FR007HF with CCD detector
Space group	$P3_1$
Unit cell parameters	
a=b (Å)	95.34
c (Å)	39.65
Resolution range (Å)	82.57–2.99
Wavelength (Å)	1.54
Average redundancy	2.7 (2.4)
Unique reflections	7669
Completeness (%)	93.6 (78.8)
Rmerge (%)	7.9 (47.9)
Average $I/\sigma(I)$	14.8 (2.1)

Values in parentheses are for the highest resolution shell (3.13–2.99 Å).

Table 3. Refinement statistics

	*Ss_Ec*Trx
Resolution range (Å)	82.57–2.99
Asymmetric unit	2 molecules
R, R_{free} (%)	19.8, 25.9
No. of residues	214
Mean B value (Å^2)	53.3
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.50

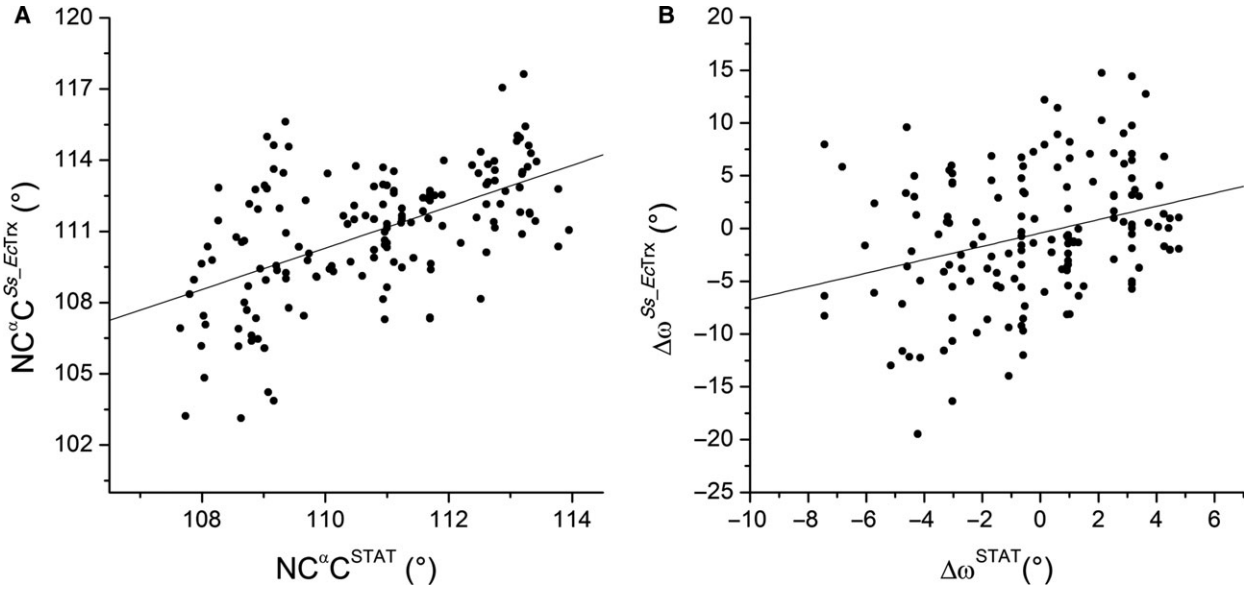

Fig. 10. Validation of the variability in some geometrical parameters of the protein backbone: the NC$^{\alpha}$C bond angle (A) and the deviation from the peptide bond planarity Δω (B) of *Ss_Ec*Trx structure (PDB ID: 6H1Y). The linear regression analysis has been performed comparing the geometrical parameters of residues with the average values obtained for residues adopting the same (φ,ψ)-conformation derived from databases of protein structures determined at high resolution (resolution <1.6 Å for NC$^{\alpha}$C and < 1.2 Å for Δω). The regression lines are shown. The values of the correlation coefficients (and *P*-values) are reported in Table 4.

Table 4. Results of the regression analysis of the variability in some geometrical parameters of Ss_EcTrx structure versus datasets of parameters derived from well-refined structures (see text and legend of Figure 10).

Geometrical parameter	Correlation coefficient	*P*-value
NC$^{\alpha}$C	0.56	$1.3*10^{-13}$
NC$^{\alpha}$C$^{\beta}$	−0.037	0.65
C$^{\beta}$C$^{\alpha}$C	0.22	$7.4*10^{-3}$
C$^{\alpha}$CO	0.22	$6.6*10^{-3}$
C$^{\alpha}$CN$_{+1}$	0.33	$3.6*10^{-5}$
OCN$_{+1}$	0.092	0.26
C$_{-1}$NC$^{\alpha}$	0.36	$4.0*10^{-6}$
Δω	0.30	$1.8*10^{-4}$
θ_C	−0.024	0.77

data, which was randomly selected, was used in R-free calculations to monitor the progress of refinement. Despite the limited resolution of the data, the electron density is well defined for most of the residues. Weak or absent electron density has been detected for terminal residues and for few solvent-exposed side chains. The final model presents R-factor and R-free values of 0.198 and 0.259, respectively. A summary of the refinement statistics is reported in Table 3. The stereochemistry indicators of the refined model are in line with those displayed by well-refined structures as highlighted by the program PROCHECK [43]. The overall quality of the structure was also assessed by analyzing the variability in some geometrical parameters of the protein backbone [44–47]. Despite the rather low resolution of the structure, the final model is able to reproduce the variability in some backbone bond angles (NC$^{\alpha}$C, C$^{\beta}$C$^{\alpha}$C, C$^{\alpha}$CO, C$^{\alpha}$CN$_{+1}$, and C$_{-1}$NC$^{\alpha}$) and the deviation from planarity of the peptide bond (Δω) as function of conformation (Fig. 10 and Table 4). The atomic coordinates of *Ss_Ec*Trx have been deposited in the PDB with the identification code 6H1Y.

Acknowledgements

The authors thank Luca De Luca, Maurizio Amendola, and Giosuè Sorrentino for technical support.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AR and LV designed the experiments. NB, GS, LE, and AR performed the experiments. All authors analyzed the data and discussed the results. L.V. wrote the paper in close collaboration with all the authors.

References

1 Dill KA, Ozkan SB, Shell MS & Weikl TR (2008) The protein folding problem. *Annu Rev Biophys* **37**, 289–316.

2 Rose GD, Fleming PJ, Banavar JR & Maritan A (2006) A backbone-based theory of protein folding. *Proc Natl Acad Sci USA* **103**, 16623–16633.
3 Dill KA & MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* **338**, 1042–1046.
4 Petrey D & Honig B (2005) Protein structure prediction: inroads to biology. *Mol Cell* **20**, 811–819.
5 Das R & Baker D (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**, 363–382.
6 Tramontano A (2004) A brighter future for protein design. *Angew Chem Int Ed Engl* **43**, 3222–3223.
7 Moult J, Fidelis K, Kryshtafovych A & Tramontano A (2011) Critical assessment of methods of protein structure prediction (CASP)–round IX. *Proteins* **79** (Suppl 10), 1–5.
8 Kim DE, Chivian D & Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526–W531.
9 Lee J, Freddolino PL & Zhang Y (2017) Ab Initio Protein Structure Prediction. In *From Protein Structure to Function with Bioinformatics* (J. Rigden D, ed.), pp. 3–35. Springer, Dordrecht.
10 Floudas CA (2007) Computational methods in protein structure prediction. *Biotechnol Bioeng* **97**, 207–213.
11 Pucci F & Rooman M (2017) Physical and molecular bases of protein thermal stability and cold adaptation. *Curr Opin Struct Biol* **42**, 117–128.
12 Razvi A & Scholtz JM (2006) Lessons in stability from thermophilic proteins. *Protein Sci* **15**, 1569–1578.
13 Petsko GA (2001) Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol* **334**, 469–478.
14 Spadaccini R, Leone S, Rega MF, Richter C & Picone D (2016) Influence of pH on the structure and stability of the sweet protein MNEI. *FEBS Lett* **590**, 3681–3689.
15 Leone S, Pica A, Merlino A, Sannino F, Temussi PA & Picone D (2016) Sweeter and stronger: enhancing sweetness and stability of the single chain monellin MNEI through molecular design. *Sci Rep* **6**, 34045.
16 Leone S & Picone D (2016) Molecular dynamics driven design of pH-stabilized mutants of MNEI, a sweet protein. *PLoS One* **11**, e0158372.
17 Bischof JC & Xiaoming HE (2005) Thermal stability of proteins. *Ann NY Acad Sci* **1066**, 12–33.
18 Teilum K, Olsen JG & Kragelund BB (2011) Protein stability, flexibility and function. *Biochim Biophys Acta* **1814**, 969–976.
19 Tokuriki N & Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* **19**, 596–604.
20 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.
21 Magliery TJ (2015) Protein stability: computation, sequence statistics, and new experimental methods. *Curr Opin Struct Biol* **33**, 161–168.
22 Berezovsky IN, Zeldovich KB & Shakhnovich EI (2007) Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput Biol* **3**, e52.
23 Bae E, Bannen RM & Phillips GN Jr (2008) Bioinformatic method for protein thermal stabilization by structural entropy optimization. *Proc Natl Acad Sci USA* **105**, 9594–9597.
24 Kumar S, Tsai CJ & Nussinov R (2000) Factors enhancing protein thermostability. *Protein Eng* **13**, 179–191.
25 Thompson MJ & Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* **290**, 595–604.
26 Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT & Baker D (2012) Principles for designing ideal protein structures. *Nature* **491**, 222–227.
27 Balasco N, Esposito L, De Simone A & Vitagliano L (2013) Role of loops connecting secondary structure elements in the stabilization of proteins isolated from thermophilic organisms. *Protein Sci* **22**, 1016–1023.
28 Powis G & Montfort WR (2001) Properties and biological activities of thioredoxins. *Annu Rev Biophys Biomol Struct* **30**, 421–455.
29 Martin JL (1995) Thioredoxin–a fold for all reasons. *Structure* **3**, 245–250.
30 Collet JF & Messens J (2010) Structure, function, and mechanism of thioredoxin proteins. *Antioxid Redox Signal* **13**, 1205–1216.
31 Stefankova P, Kollarova M & Barak I (2005) Thioredoxin – structural and functional complexity. *Gen Physiol Biophys* **24**, 3–11.
32 Canali E, Bolchi A, Spagnoli G, Seitz H, Rubio I, Pertinhez TA, Muller M & Ottonello S (2014) A high-performance thioredoxin-based scaffold for peptide immunogen construction: proof-of-concept testing with a human papillomavirus epitope. *Sci Rep* **4**, 4729.
33 Brown CJ, Dastidar SG, See HY, Coomber DW, Ortiz-Lombardia M, Verma C & Lane DP (2010) Rational design and biophysical characterization of thioredoxin-based aptamers: insights into peptide grafting. *J Mol Biol* **395**, 871–883.
34 Moretto N, Bolchi A, Rivetti C, Imbimbo BP, Villetti G, Pietrini V, Polonelli L, Del Signore S, Smith KM, Ferrante RJ *et al.* (2007) Conformation-sensitive antibodies against Alzheimer amyloid-beta by immunization with a thioredoxin-constrained B-cell epitope peptide. *J Biol Chem* **282**, 11436–11445.
35 Leone M, Di Lello P, Ohlenschlager O, Pedone EM, Bartolucci S, Rossi M, Di Blasio B, Pedone C, Saviano M, Isernia C *et al.* (2004) Solution structure and backbone dynamics of the K18G/R82E Alicyclobacillus

acidocaldarius thioredoxin mutant: a molecular analysis of its reduced thermal stability. *Biochemistry* **43**, 6043–6058.
36 Pedone E, Bartolucci S, Rossi M, Pierfederici FM, Scire A, Cacciamani T & Tanfani F (2003) Structural and thermal stability analysis of Escherichia coli and Alicyclobacillus acidocaldarius thioredoxin revealed a molten globule-like state in thermal denaturation pathway of the proteins: an infrared spectroscopic study. *Biochem J* **373**, 875–883.
37 Ruggiero A, Masullo M, Marasco D, Ruocco MR, Grimaldi P, Arcari P, Zagari A & Vitagliano L (2009) The dimeric structure of *Sulfolobus solfataricus* thioredoxin A2 and the basis of its thermostability. *Proteins* **77**, 1004–1008.
38 Esposito L, Ruggiero A, Masullo M, Ruocco MR, Lamberti A, Arcari P, Zagari A & Vitagliano L (2012) Crystallographic and spectroscopic characterizations of *Sulfolobus solfataricus* TrxA1 provide insights into the determinants of thioredoxin fold stability. *J Struct Biol* **177**, 506–512.
39 Roos G, Garcia-Pino A, Van Belle K, Brosens E, Wahni K, Vandenbussche G, Wyns L, Loris R & Messens J (2007) The conserved active site proline determines the reducing power of *Staphylococcus aureus* thioredoxin. *J Mol Biol* **368**, 800–811.
40 Mossner E, Huber-Wunderlich M & Glockshuber R (1998) Characterization of Escherichia coli thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. *Protein Sci* **7**, 1233–1244.
41 Manning MC, Chou DK, Murphy BM, Payne RW & Katayama DS (2010) Stability of protein pharmaceuticals: an update. *Pharm Res* **27**, 544–575.
42 Jancarik J, Scott WG, Milligan DL, Koshland DE Jr & Kim SH (1991) Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium. *J Mol Biol* **221**, 31–34.
43 Laskowski RA, Macarthur MW, Moss DS & Thornton JM (1993) Procheck – a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
44 Balasco N, Esposito L, Thind AS, Guarracino MR & Vitagliano L (2017) Dissection of factors affecting the variability of the peptide bond geometry and planarity. *Biomed Res Int* **2017**, 2617629.
45 Balasco N, Esposito L & Vitagliano L (2017) Factors affecting the amplitude of the tau angle in proteins: a revisitation. *Acta Crystallogr D Struct Biol* **73**, 618–625.
46 Improta R, Vitagliano L & Esposito L (2011) Peptide bond distortions from planarity: new insights from quantum mechanical calculations and peptide/protein crystal structures. *PLoS One* **6**, e24533.
47 Esposito L, De Simone A, Zagari A & Vitagliano L (2005) Correlation between omega and psi dihedral angles in protein structures. *J Mol Biol* **347**, 483–487.