Thioredoxin from *Bacillus acidocaldarius*: characterization, high-level expression in *Escherichia coli* and molecular modelling

Simonetta BARTOLUCCI^{*1}, Annamaria GUAGLIARDI^{*}, Emilia PEDONE^{*}, Donatella DE PASCALE^{*}, Raffaele CANNIO^{*}, Laura CAMARDELLA[†], Mosè ROSSI[‡], Giuseppe NICASTRO[‡], Cesira de CHIARA[‡], Paolo FACCI[‡], Giancarlo MASCETTI[‡] and Claudio NICOLINI[‡]

^{*}Dipartimento di Chimica Organica e Biologica, Università 'Federico II' di Napoli, Via Mezzocannone 16, Napoli 80134, Italy, [†]Istituto di Biochimica delle Proteine ed Enzimologia, C.N.R., Via Marconi 10, Napoli 80125, Italy, and [‡]Istituto di Biofisica, Università di Genova, Via Giotto 2, Genova Sestri Ponente 16153, Italy

The thioredoxin (Trx) from *Bacillus acidocaldarius* (BacTrx) was purified to homogeneity by anion-exchange chromatography and gel-filtration chromatography, based on its ability to catalyse the dithiothreitol-dependent reduction of bovine insulin disulphides. The protein has a molecular mass of 11 577 Da, determined by electrospray mass spectrometry, a pI of 4.2, and its primary structure was obtained by automated Edman degradation after cleavage with trypsin and cyanogen bromide. The sequences of known bacterial Trxs were aligned at the active site: BacTrx has an identity ranging from 45 to 53 % with all sequences except that of the unusual *Anabaena* strain 7120 Trx (37 % identity). The gene coding for BacTrx was isolated by a strategy based on PCR gene amplification and cloned in a plasmid downstream of a *lac*-derived promoter sequence; the recombinant clone was used as the expression vector for *Escherichia coli*. The expression was optimized by varying both the time of cell growth

and the time of exposure to the inducer isopropyl β -D-thiogalactoside; expressed BacTrx represents approx. 5 % of the total cytosolic protein. CD spectra and differential scanning calorimetry measurements demonstrated that BacTrx is endowed with a higher conformational heat stability than the Trx from *E. coli*. Nanogravimetry experiments showed a lower content of bound water in BacTrx than in *E. coli* Trx, and a transition temperature approx. 10 °C higher for BacTrx. The three-dimensional model of the oxidized form of BacTrx was constructed by a comparative molecular modelling technique, using *E. coli* Trx and *Anabaena* strain 7120 Trx as reference proteins. Increased networks of ion-pairs and shorter loops emerged as major features of the BacTrx structure compared with those of the template proteins. The findings are discussed in the light of the current knowledge about molecular determinants of protein stability.

INTRODUCTION

Thioredoxin (Trx) is a protein disulphide oxidoreductase, typically with a molecular mass of 10–12 kDa and an acidic pI, and is present in eukaryotic and prokaryotic cells (for reviews see [1–3]). The protein has at the N-terminus a redox-active disulphide/dithiol in the conserved sequence -Trp-Cys-Gly-Pro-Cys-, with the only exception represented by the Trx C-2 of *Corynebacterium nephridii*, where Ala replaces Gly. Trx plays a key role in different fundamental cellular processes [4], such as deoxyribonucleotide biosynthesis, regulation of photosynthetic events, regeneration of oxidative damage, and activation of transcription factors.

A large body of literature focuses on the structural features of Trx (the 'thioredoxin fold'). Trx of *Escherichia coli* represents the reference model for these investigations, since the structure in solution of the oxidized and the reduced forms has been calculated by NMR [5], and the crystal structure of the oxidized form solved [6]. The molecule has a central core of five strands of β -sheet enclosed by four α -helices, with the redox-active sequence in a reverse turn in a β - α - β structural unit; this organization confers an excellent stability to the molecule.

The important role played by Trx in cell functionality has recently opened the door to its technological application to the

solution of a number of problems in nutrition and medicine, such as food allergenicity, inflammatory diseases, and progression of viral diseases [4]. Trx has been proven to be useful as a purification tag for the overexpression of economically valuable proteins and peptides [7]. More recently, it has been shown that the solubility of foreign proteins in *E. coli* increases by co-production of Trx [8].

This paper concerns the Trx of *Bacillus acidocaldarius* (BacTrx), the first Trx so far described from a thermophilic source. The physicochemical features of the protein, its primary structure, and the high-level expression in *E. coli* of the coding gene are reported. The structural stability of BacTrx has been investigated and compared with that of the Trx of *E. coli* by means of CD, differential scanning calorimetry (DSC) and nanogravimetry. The molecular modelling of oxidized BacTrx allowed the identification of the possible molecular determinant(s) of its high heat stability.

MATERIALS AND METHODS

Materials

Bovine insulin and 4-vinylpyridine were from Sigma. Molecular-mass standards for SDS/PAGE were obtained from Gibco BRL.

Abbreviations used: BacTrx, the thioredoxin from *Bacillus acidocaldarius*; DSC, differential scanning calorimetry; DTT, dithiothreitol; IPTG, isopropyl β -D-thiogalactoside; TFA, trifluoroacetic acid; Trx, thioredoxin.

¹ To whom correspondence should be addressed.

The novel nucleotide sequence data published here have been deposited with the EMBL/GenBank/DBJ nucleotide sequence data banks and are available under the accession number P80579.

Ampholynes, pTrc 99A and pUC 18 *Sma*I/BAP plasmids were from Pharmacia. Radioactive materials were obtained from Amersham International. Deoxynucleotides and restriction and modification enzymes were purchased from Boehringer. *E. coli* strain Rb791 was kindly provided by Professor G. Sannia (Dipartimento di Chimica Organica e Biologica, Università di Napoli, Italy). All materials used for gene amplification were supplied by Stratagene Cloning Systems. All synthetic oligonucleotides were purchased from PRIMM (Milan, Italy). *E. coli* Trx in oxidized form was purchased from IMCO [3], and was homogeneous on SDS/PAGE, salt free and used without further manipulations. All other chemicals were of the highest grade available.

Analytical methods for protein and DNA

Protein concentration was determined by the Pierce method [9] using BSA as the standard. SDS/PAGE was performed according to Laemmli [10]: samples were heated at 100 °C for 30 min in 2% (w/v) SDS/5% (v/v) 2-mercaptoethanol, and the run was performed using 5% (w/v) stacking gel and 15% (w/v) separating gel. Isoelectric focusing was carried out on a 4.6% (w/v) polyacrylamide gel at 4 °C using 0.8% (v/v) ampholytes in the pH range 2.5–7; the pI was estimated using a surface electrode [11]. After the run, the bands were revealed by the silver procedure described by Rabilloud et al. [12].

N-terminal sequences were performed by Edman degradation on pulsed liquid-phase sequencer model 477A (Applied Biosystems) equipped with a 120A analyser for on-line detection of phenylthiohydantoin amino acids.

DNA electrophoresis on 1% (w/v) agarose gels in TBE buffer (90 mM Tris/borate/20 mM EDTA), and plasmid transformations of *E. coli* cells were carried out as described by Sambrook et al. [13]. DNA sequencing was carried out by the dideoxy chain termination method [14] with [³⁵S]dATP, using the Sequenase version 2.0 sequencing kit (Amersham) on alkali-denatured double-stranded templates; the universal reverse primer and specific synthetic oligonucleotides were used. All DNA sequences were determined in duplicate on both strands. Gene sequence analysis were performed using the Clustall, Palign and Nalign PC-GENE programs (Intelligenetics, Mountain View, CA, U.S.A.).

Activity assay and protein purification

The insulin reductase activity was assayed according to Holmgren [15]. The standard assay mixture contained 0.1 M sodium phosphate buffer (pH 7.0)/0.13 mM bovine insulin, in the absence or in the presence of the protein (final volume 1 ml); upon the addition of 1 mM dithiothreitol (DTT), the increase in attenuation at 650 nm was monitored at 30 °C.

Buffers degassed and flushed with N₂ were used throughout the purification manipulations. Samples were concentrated by ultrafiltration through a membrane of 1000 Da cut-off using an Amicon cell. The insulin precipitation assay was used to monitor the protein activity.

Preparation of the crude extract

B. acidocaldarius was grown at 55 °C and pH 3.0 in a medium containing 0.5% (w/v) glucose/0.2% (w/v) yeast extract/0.01% (w/v) MgSO₄/0.01% (w/v) KH₂PO₄. Cells, harvested after 11 h, were immediately frozen in liquid N₂ and stored at –80 °C until required.

Bacteria (30 g) underwent freeze–thawing 3 times, were mixed

with 30 g of sand and 30 ml of 20 mM Tris/HCl buffer (pH 8.4)/2 mM EDTA/1 M NaCl, and homogenized in a Homni Mixer. The homogenate was centrifuged at 4000 g for 20 min at 4 °C to remove the sand, and the supernatant was sonicated 3 times with a frequency of 0.9 Hz and ultracentrifuged at 160 000 g for 90 min at 4 °C. The supernatant of the ultracentrifugation constituted the crude extract.

Anion-exchange chromatography

The crude extract (700 mg) was extensively dialysed against 20 mM Tris/HCl buffer (pH 8.4)/2 mM EDTA (Buffer A) and loaded on to a DEAE–Sephacrose Fast Flow column (Pharmacia, 2.2 cm × 18 cm) equilibrated in Buffer A. Bound proteins were eluted by a linear gradient of 0–0.2 M NaCl in Buffer A (200 ml) at a flow rate of 50 ml/h. The active fractions (2.5 ml each) were pooled and concentrated.

Gel-filtration chromatography

The DEAE pool (50 mg) was loaded in two separate runs on to a HiLoad Superdex 75 column (Pharmacia, 2.6 cm × 60 cm) connected to an FPLC system (Pharmacia), eluted with 20 mM Tris/HCl buffer (pH 8.4)/0.2 M NaCl (Buffer B) at a flow rate of 2 ml/min. The active fractions were pooled and concentrated (3 mg).

Protein alkylation and cleavage

The freeze-dried protein (1 mg) was dissolved in 0.5 M Tris/HCl (pH 8.4)/2 mM EDTA/6 M guanidinium chloride; DTT was added at a 5-fold molar excess over cysteine residues, and the solution was incubated at 30 °C for 3 h under N₂. Alkylation was obtained by a 1 h reaction with 4-vinylpyridine (100-fold molar excess over cysteine residues) in the dark at room temperature under N₂. The protein was desalted by gel filtration on a prepacked G-25 column (PD-10, Pharmacia) equilibrated with 0.1% (v/v) trifluoroacetic acid (TFA), and it was freeze-dried. The *S*-pyridyl-ethylated protein was resuspended in 1% (w/v) ammonium bicarbonate and digested with trypsin added in two equal portions to a final ratio of 1:50 (w/w) for 4 h at 37 °C. The peptide mixture was freeze-dried, resuspended in 0.1% TFA and loaded on to a reversed-phase C18 μ Bondapak column (Waters, 0.39 cm × 30 cm) connected to a HPLC System Gold (Beckman). Peptides were eluted with a linear gradient of 5–60% (v/v) acetonitrile in 0.1% TFA, and detected by monitoring the absorbance at 220 nm and 280 nm. CNBr cleavage was obtained by incubating the *S*-pyridyl-ethylated protein in 75% (v/v) formic acid with CNBr (100-fold molar excess over the estimated methionine concentration) for 24 h at room temperature. Excess CNBr was removed by speedy vacuum centrifugation; the peptides were separated by SDS/PAGE as described, electroblotted on a Problott membrane (Applied Biosystems) according to the manufacturer's method, and were sequenced.

High-level expression in *E. coli* and purification of the recombinant protein

Isolation of chromosomal DNA from *B. acidocaldarius* was performed as described by Barker [16].

Based on the BacTrx amino acid sequence from residues 1–7 and that from residues 96–107, the following oligonucleotides were designed and used as primers in the PCR gene amplification procedure, using the chromosomal DNA (10 ng) as template:

5'-GCCCATGGCTACGATGACGTTGACTGATG-3' (29mer) and 5'-GATTACTGTAATACATCTGCTAACTGTGCCTC-CAAC-3' (36mer). The amplification was performed according to Saiki [17] for 35 cycles at 55 °C annealing temperature, on a Perkin Elmer Cetus Cycler Temp. The amplified DNA fragment (*BacTrx*) was made blunt by filling-in with the Klenow fragment, dCTP, dGTP, dATP and dTTP. *BacTrx* was cloned in pUC 18 *Sma*I/BAP and sequenced. The plasmid containing the gene was digested with *Nco*I and *Kpn*I, and it was inserted in the pTrc 99A plasmid previously digested with the same enzymes. The recombinant clone, designated pTrx, represented the expression vector.

E. coli Rb791 competent cells were transformed with pTrx and grown at 37 °C to different cell densities in 1 litre of Luria–Bertani medium; isopropyl β -D-thiogalactoside (IPTG) was added to 1 mM final concentration, and the induction time was varied from 2 to 24 h. Rb791 cells transformed with pTrc 99A represented a negative control. Cells were harvested by centrifugation at 3000 g for 10 min at 4 °C, and washed with 30 ml of ice-cold 20 mM Tris/HCl buffer (pH 8.4)/0.1 M NaCl/1 mM EDTA. Cells were disrupted by a 20 min incubation at room temperature in the presence of lysozyme (4 mg/ml), added with 70 ml of the same buffer supplemented with PMSF (0.7 mM). After the addition of sodium deoxycholate (2.4 mM) and DNase I (10 nM), the suspension was incubated at room temperature for 30 min, and then centrifuged at 5000 g for 15 min at 4 °C. The supernatant of a centrifugation at 5000 g for 15 min at 4 °C represented the crude extract. Starting from the crude extract of transformed cells grown at 1.0 D_{600} cell density, and induced for 18 h with IPTG, the recombinant protein was purified to homogeneity by means of two successive heat treatments (at 80 °C for 20 min), an anion-exchange chromatography and a gel-filtration chromatography performed as described for the native protein.

Spectroscopic techniques and DSC

The molecular mass of the *S*-pyridyl-ethylated protein was measured on a Hewlett Packard 5998B quadrupole mass spectrometer, equipped with a 59987A electrospray source. The protein, dissolved in 50 % (v/v) methanol/1 % (v/v) acetic acid, was injected by means of a syringe pump at a flow rate of 2 μ l/min.

Fluorescence spectra were recorded at 25 °C with a thermostatically controlled Perkin Elmer LS-50B spectrofluorimeter; samples were excited at 295 nm, and the emission spectra were recorded from 310 to 440 nm. The protein (40–80 μ g/ml) was dissolved in 50 mM sodium phosphate, pH 7.0, and reduced by 10 min incubation in 2.5 mM DTT at 25 °C. The effects of pH were determined in the following buffers with an ionic strength of 50 mM: sodium citrate (pH 3.0–4.6); sodium acetate (pH 4.5–5.5); sodium phosphate (pH 5.5–7.5); Tris/HCl (pH 8.0–9.0); glycine/NaOH (pH 9.5–10.5).

CD spectra were recorded on a Jasco J-710 spectropolarimeter equipped with a Peltier thermostatic cell holder (Jasco, model PTC-343). Samples were at a concentration of 0.2 mM (*BacTrx*) and 0.085 mM (*E. coli* Trx) in 10 mM sodium phosphate, pH 5.8. Molar ellipticity per residue, $[\theta]$, expressed in deg $\cdot$ cm² $\cdot$ dmol⁻¹, was calculated from the formula $[\theta] = MW \theta / (100cn)$, where θ is the ellipticity observed (m $\cdot$ deg), MW is the molecular mass, c the protein concentration (mol/dm⁻³), l is the path length of the cuvette (cm) and n is the number of amino acid residues.

DSC measurements were performed on a DSC-2C (Perkin Elmer) at a rate of 60 °C $\cdot$ h⁻¹ in duplicate for each sample, at a

concentration of 23.3 mg/ml (*BacTrx*) and 23.1 mg/ml (*E. coli* Trx), in 50 mM sodium phosphate, pH 5.8.

Nanogravimetry

Nanogravimetry exploits the properties of piezoelectric quartz crystals to vary their resonance frequency (f_0) when a mass (Δm) is adsorbed to or desorbed from their surface according to the equation [18,19]:

$$\Delta f/f_0 = -\Delta m/A\rho l$$

where A is the area covered by the adsorbate and ρ and l are the quartz density and thickness respectively.

The technique is used for measuring small inertial masses, in the range from 1 ng to several mg, deposited or adsorbed on to the surface of the transducer due to (bio)chemical reactions or physical adsorption. It has also been used for estimating the amount of water present in dry smears [20] or Langmuir–Blodgett films of proteins [21], and for obtaining isotherms for biological materials. It has been shown that maxima in the desorption rate take place in correspondence to the unfolding of proteins, when the protein globule is suddenly broken and non-covalent forces maintaining the bound water can no longer prevail against the increase in entropy [21]. Therefore, the technique provides information on the transition temperature of protein unfolding under the chosen experimental conditions.

Quartz resonators (10 MHz) with aluminium electrodes with a 0.57 ± 0.18 ng/Hz sensitivity and a 0.3 ± 0.1 ng detection limit were used. The sample (10 μ l of 0.2 mg/ml protein solution) was deposited on the quartz crystal and dried under vacuum for 30 min. After measuring the absolute initial water amount, the temperature was increased (from 25 to 150 °C) for 5 min. The quartz was kept at room temperature, and the desorbed mass amount was measured from the frequency shift.

Molecular modelling

The atomic co-ordinates of the crystallographic structures at 1.68 Å resolution of the oxidized forms of *E. coli* (monomer A) Trx [6] and *Anabaena* strain 7120 Trx [22] were obtained from the Protein Data Bank files [23]. The alignment of *BacTrx* to its templates was based on a multiple sequence alignment, performed with the program MaxHom [24], and the structure was constructed by the program MODELLER [25], based on satisfaction of spatial restraints without further manual intervention. The molecular graphics program InsightII (BIOSYM Software) was used on a Silicon Iris Crimson workstation. Out of the five models derived, the representative model is that having the lowest value of the objective function, and its quality has been evaluated by means of PROCHECK [26], for stereochemical analysis, and WHAT IF [27], for quality control option.

The quality of the deduced model was verified by appropriate programs; the stereochemical parameters are consistent with a model of 1.8 Å resolution, and the model has a distribution of Φ , ϕ angles, which is similar to that expected for experimental structures. A Ramachandran plot of the modelled protein shows that 97.8 % of residues are found in most favoured regions, and that most of the rotamers in the model are localized in low-energy regions. Upon superposition on to template proteins, *BacTrx* (residues 3–103) exhibited a root-mean-square deviation of 1.1 Å and 1.23 Å respectively for C- α atoms, this confirming its very high similarity to the template proteins.

Ion-pair interactions were identified using the criterion of a distance ≤ 4 Å between charged groups [28]. The secondary structures were defined by the main-chain hydrogen bonding pattern using the DSSP program [29].

RESULTS AND DISCUSSION

Protein purification, pI, molecular mass and primary structure

The reduction of protein disulphides catalysed by reduced Trx can easily be measured using insulin as substrate: upon reduction of the inter-chain disulphide bridges between chains A and B of insulin, the turbidity of the solution increases due to precipitation of the free B chain [15]. All Trxs so far reported are effective in this assay.

The crude extract of *B. acidocaldarius* was able to catalyse the DTT-dependent reduction of bovine insulin disulphides. Exploiting the knowledge that Trxs are acidic proteins, we applied the crude extract of *B. acidocaldarius* to a DEAE-Sephacel column, performed at pH 8.4: no insulin reductase activity was found in the flow-through of the column, and a sharp peak of activity was eluted by a linear gradient of ionic strength (centred at 0.18 M NaCl). The active sample from the anion-exchange chromatography was fractionated on the basis of the molecular size on a Superdex 75 column: the peak of activity had an elution volume that corresponded to 10 kDa in a calibrated run. SDS/PAGE analysis of the sample from the gel-filtration column showed one single band whose molecular mass was approx. 10 kDa (Figure 1). Isoelectric-focusing gel electrophoresis of the protein solution revealed one band with a pI of 4.2.

Figure 2 shows the DTT-dependent reduction of insulin disulphides at 30 °C in the presence of increasing concentrations of pure BacTrx, and the spontaneous precipitation reaction. The progress curve obtained for 10 µM protein (not shown) was identical with that for 8 µM, indicating that the assay was saturated. The insulin reductase activity of BacTrx was unaffected by previous heating of the protein to 80 °C for 7 h or to 90 °C for 1 h. After 3 h of incubation at 90 °C, BacTrx gave a time-to-precipitation of 9 min, versus 7.5 min for the non-heated protein; the protein was completely inactive after 7 h of incubation at 90 °C.

The molecular mass of the protein was determined by electrospray mass spectrometry; a value of 11 787 Da was calculated, which corresponded to the mass deduced from the amino acid sequence (see below) plus the two pyridyl-ethyl moieties (210 Da) present at the cysteine residues.

The complete amino acid sequence of BacTrx was determined by direct Edman degradation of the entire protein and by sequencing the peptides derived from trypsin digestion; the

Figure 2 Insulin reductase activity of BacTrx

The DTT-dependent reduction of bovine insulin disulphides was carried out as described in the Materials and methods section in the absence (○) or in the presence of different concentrations of pure BacTrx: (●) 1 µM; (■) 4 µM; (▲) 8 µM.

peptide obtained by CNBr cleavage was partially sequenced after electroblotting from an SDS/PAGE gel (Figure 3). Fragments T1, T2 and T3 were aligned by means of the overlapping N-terminal sequence and the sequence of the CNBr peptide; T6 was located at the C-terminus because of the absence of a Lys or Arg residue; T4 and T5 were aligned by similarity with other Trx sequences (their position was confirmed by sequencing of the gene). The total sequence comprises 105 amino acid residues, corresponding to a calculated mass of 11 577 Da (in the reduced form).

The primary structure of BacTrx was compared with those of all bacterial Trxs so far reported; sequences were aligned from the active site, and gaps were introduced to maximize similarity (Figure 4). Major sites of sequence identity among the proteins can be identified (numbering and structural elements of *E. coli* Trx): the region preceding the active site (residues 22–29); the active site (residues 31–37); the regions corresponding to strand β 3 and helix α 3 (residues 52–64); and the region corresponding to strand β 4 (residues 76–84). The identity is low at the N- and C-termini. BacTrx has 49% identity with *E. coli* Trx. The percentage of identity ranges from 45 to 53% with all sequences, except that of the unusual Trx from *Anabaena* strain 7120 that shares 37% identity with BacTrx.

High-level expression in *E. coli*

Trx is a protein that is present in amounts that often pose a limit to its characterization and use. The size of BacTrx suggested a rapid procedure to obtain milligram amounts by cloning and expressing the coding gene in *E. coli*. A PCR amplification strategy, using specific oligonucleotides, designed on the basis of protein sequence and *B. acidocaldarius* codon usage, allowed the isolation of the coding sequence already adapted to be inserted into the expression vector. Moreover, this strategy avoided the classical procedures of library constitution and screening for gene cloning.

The newly constructed plasmid pTrx was used as expression vector: in fact, the transcription of recombinant genes is controlled by the strong hybrid regulatory sequence promoter ptac, derived from the fusion of the *lac* and *trp* promoters, and is still inducible by IPTG. *E. coli* Rb791 competent cells were transformed with pTrx, and allowed to grow in aliquots up to 0.5, 1, 1.5, 2 and 2.5 D_{600} cell densities. The overexpression of BacTrx

Figure 1 SDS/PAGE analysis of the active sample during the purification procedure of BacTrx

Lane 1, crude extract; lane 2, pooled active fractions from DEAE-Sephacel column; lane 3, pooled active fractions from Superdex 75 column; lane 4, molecular-mass standards with masses indicated on the right.

Figure 3 Amino acid sequence of BacTrx

Details are in the Materials and methods section. Tryptic peptides (T) are underlined; the continuous arrow indicates the residues determined by Edman degradation of the entire protein; the broken arrow indicates the peptide derived from CNBr cleavage.

	<u>β1</u>	<u>α1</u>	<u>β2</u>	<u>α2</u>	<u>β3</u>	<u>α3</u>	
<i>E. coli</i>	SDKII H LTDD S	FD T DVL KA	DG	AILVDFWAEWCGPCRM I	APILDEIA	DE	YQ G KLT VAKLNID QN
<i>BacTrx</i>	ATM TLTD A	NF QQA I	QGD K	PVLVDFWAEWCGPCRM	MAPVLEEF A	E A H A D	KVT VAKLNIDE N
<i>C. nephridii C-2</i>	S ATIVNT TDE	NF Q ADVL	D	AETPVLNDFWAGWCA	PKA I	APVLEEL SN	E Y AG KV KIV K VDVTS
<i>C. nephridii C-1</i>	AT V KV D NSNF QS	DVL Q	SSE	PVVVDFWAEWCGPCRM	I	APALDEIAT	EMA G Q V KI AKVNI
<i>Anabaena 7119</i>	S AAA Q VTD	STFKQ EVL	DS	D	VPVLVDFWAEWCGPCRM	V	APVVDEIAQQ Y EG KI KVV KVNIDE N
<i>Anabaena 7120</i>	SKGVI TITDAE	FE S EVL KAEQ		PVLVYFWASWCGPC	QLMSP L I	NL	AANTY SD RL KVV KLEIDP N
<i>A. nidulans R2</i>	MSVAA A VTDA	TFKQ EVL	ESS	IPVLVDFWAEWCGPCRM	V	APVVDEIAQQ	Y SDQVKV V KVNIDE N
<i>C. thiosulfatophilum</i>	AGKYFEATDK	NF QT EXX	DS	D	K AVXVDFWAEWCGPC	MMXG PVX E	QX ADDY EG KA XXAKXNIDE N
<i>C. vinosum</i>	SDSI VH VTDD S	FEE EVX K	SPD	PVLVDYWADWCGPCRM	X	APVXDEIA	DE Y AG RV KX AKXNIDE N
<i>R. sphaeroides</i>	S T VP VTDA	TFD T EV RK	S D	VPVVDFWAEWCGPCR	QIG	PALEEL S	KE Y AG KV KIV KVNID QN
<i>R. rubrum</i>	M K Q VSDA S	FEE DVL KA	DG	PVXVDFWAEWCGPCR	QX	APALEELAT	AL GD KVT VAKINIDE N
<i>B. subtilis</i>	M A IVK ATDQ S	F SAE	TSEG	V	VLADFWEWCGPCRM	I	APVLEEL D QEMGD KL KIV KIDVIDE N
<i>T. ferrooxidans</i>	MSDAILY VSDD S	FE T DVL K	SSK	PVLVDFWAEWCGPCRM	I	APILEEIA	DE Y A DRLR VAKENIDE N
<i>S. typhimurium</i>	MSDKII H LTDD S	FD T DVL KA	DG	AILVDFWAEWCGPCRM	I	APILDEIA	DE YQ G KLT VAKLNID QN
consensus	tD	F v		pvlvDFWA WCGPC m	aP l E		y g k v K n D N

	<u>β4</u>	<u>β5</u>	<u>α4</u>	
<i>E. coli</i>	G T APKY GIRGIP T	LL LFKNGE VA AT	K VGAL S KGQLKE F L D	A N L A (108)
<i>BacTrx</i>	ETTSQF GIMSIP LT	IL FKGGRPV	KQLIGY QPKQL EA Q LAD V	LQ (105)
<i>C. nephridii C-2</i>	EDTA V KYN IRNIPA	LLM FKDGEEVA Q	Q VGAAPRS K LA AF I D	Q N I (107)
<i>C. nephridii C-1</i>	ELAAQF GVR SIP T	LLM FKDGE LAA N	M VGA A PK SRL A	DW IKAS A (105)
<i>Anabaena 7119</i>	Q VASQ Y GIRSIP T	LMIFKGGQKV D	MVVG A V PKTTL S	QTL E K H L (106)
<i>Anabaena 7120</i>	TTV K KYK VEGVPALR	LV K GEQILDST E	G VI S KDKLLS F L D T	HLN NN (110)
<i>A. nidulans R2</i>	S VASQ Y GIRSIP T	LMIFKDGQVR D	T VVGA V PKTTL A	NTL D K H L (107)
<i>C. thiosulfatophilum</i>	N XAGQ Y GXRSXP	TM XXX KGGKV D	QMVGA X PKNMX A	KKX D E H IG (108)
<i>C. vinosum</i>	N TPP R Y GXRGIP T	LMLFRGGE V	EAT K VGA V S KSQL	TAF L D S N X (107)
<i>R. sphaeroides</i>	E S PAMLGVRGIPALF	L FKNQGVV SN	K VGA A PKAAL A	T W I AS A L (105)
<i>R. rubrum</i>	Q TPS KY GVRGIP T	LMIFKDGQ VA	AT K IGA L PKTKL F	EW VEAS V (104)
<i>B. subtilis</i>	QETAG KY GVMSIP T	LLVL KDGEVV E	TS VGF K PKEAL Q	E L NKH L (104)
<i>T. ferrooxidans</i>	N TPPQYAIR	GIP T	LLL FKAGKL E T K	VGA L S KAQL TAF L DS Q L (108)
<i>S. typhimurium</i>	G T APKY GIRGIP T	LLL FKNGE VA	AT K VGA L S KGQLKE F L D	A N L A (109)
consensus	y g r i P t	L fK G v	vGa K L	d

Figure 4 Comparison of the sequences of bacterial Trxs

The alignment was performed as described in the text. The sequences were from the following sources: *E. coli* [34]; BacTrx (this work); *Corynebacterium nephridii* C-2 [35]; *C. nephridii* C-1 [36]; *Anabaena* strain 7119 [37]; *Anabaena* strain 7120 [38]; *Anacystis nidulans* R2 [39]; *Chlorobium thiosulfatophilum* [40]; *Chromatium vinosum* [41]; *Rhodobacter sphaeroides* [42]; *Rhodospirillum rubrum* [43]; *Bacillus subtilis* (PIR Release Accession No. B37192); *Thiobacillus ferrooxidans* (GenPept Accession No. U20361); *Salmonella typhimurium* (PIR Release Accession No. S35497). Positional numbers refer to the sequence of the *E. coli* protein. The length of each sequence is indicated in parentheses. X denotes either leucine or isoleucine. The residues in bold type are identical with the sequence of BacTrx. The consensus for all sequences (capital letter), for all but one sequence (capital letter underlined), and for at least ten out of fourteen sequences (lower-case letter) is reported. The consensus at positions 79 and 99 is based on homology considerations.

was induced by the addition of 1 mM IPTG to the culture media, the induction time being 16 h in each case. The five extracts, and the extract from *E. coli* cells transformed with the expression vector lacking any inserted sequence (the negative control), were assayed for the insulin reductase activity using the same protein

amount: maximum activity was found in the extract from transformed cells grown at 1 D_{600} cell density. To further optimize the expression of the protein, transformed cells grown at 1 D_{600} cell density were exposed to 1 mM IPTG for increasing time lengths, and the insulin reductase activity was assayed on each

Figure 5 CD spectra for BacTrx (panel A) recorded at (a) 25 °C, (b) 74 °C, (c) 80 °C, (d) 86 °C and (e) 88 °C, and for *E. coli* Trx (panel B) recorded at (a) 25 °C, (b) 77 °C, (c) 81 °C, (d) 86 °C and (e) 88 °C

Ordinate, $10^{-3} \times [\theta]$ (deg · cm² · dmol⁻¹).

extract as described: the results showed that the maximum level of expression was reached after an 18 h induction.

For preparative purposes, the crude extract from transformed cells was grown at 1 D_{600} cell density and induced for 18 h with IPTG. The extract of *E. coli* was subjected to two thermal precipitation steps, exploiting the known stability of Trxs; an anion-exchange chromatography and gel-filtration chromatography yielded a solution that revealed only one protein band on performing SDS/PAGE analysis (not shown). The recombinant protein represented 5.4% of total cytosolic amount; it showed pI, catalytic activity, molecular mass upon electrospray mass spectrometry analysis and primary structure identical with those of native BacTrx.

Structural investigations by spectroscopy, calorimetry and nanogravimetry

Oxidation/reduction of most Trxs results in changes in the fluorescence spectrum, due to the proximity of tryptophan residues to the redox sequence of the active site. The fluorescence emission spectra of the oxidized and the reduced forms of BacTrx [BacTrx-S₂ and BacTrx-(SH)₂ respectively], at neutral pH and 25 °C, revealed a single maximum at 350 nm for both forms, the intensity for BacTrx-(SH)₂ being 3-fold higher than

Figure 6 DSC rough data for BacTrx (panel A) and *E. coli* Trx (panel B)

a.u., arbitrary units.

that of BacTrx-S₂. At pH values varying from 3 to 10.5, no changes in tryptophanyl fluorescence intensity were observed for BacTrx-S₂, whereas a pH dependence was detected for BacTrx-(SH)₂, with a maximum at pH 5.0; both forms showed a red shift of emission maximum at pH 2.5 and from 9.0 to 10.5, which indicates protein instability at these pH values. On the whole, the results of fluorescence measurements obtained for BacTrx are similar to those reported for *E. coli* Trx [30], as expected from the high identity of their active-site regions.

Figure 5 reports the CD spectra in the far-UV region of BacTrx (panel A) and *E. coli* Trx (panel B) as a function of temperature in the range from 25 to 88 °C. *E. coli* Trx underwent significant structural variations upon increasing the temperature, with a substantial change occurring at approx. 81 °C; the spectrum at 88 °C was remarkably altered in both positive and negative signals. Spectra of BacTrx revealed a better structural preservation. The spectrum at 88 °C showed just a small decrease in ellipticity concerning almost exclusively the positive signal at 196 nm.

E. coli Trx and BacTrx were subjected to DSC measurements. The transition temperature for BacTrx (Figure 6, panel A) was 103 °C, approx. 15 °C higher than the value for *E. coli* Trx, as calculated in the present work (Figure 6, panel B) and as reported in the literature [31]. The calculated values of enthalpic

Figure 7 Water desorption spectrum for BacTrx (solid line) and *E. coli* Trx (broken line)

The water amount has been normalized to the initial protein amount in the sample (panel A). Panel B shows the water desorption rate for BacTrx (solid line) and *E. coli* Trx (broken line).

variation, ΔH , were 61.36 kcal/mol for BacTrx and 39.89 kcal/mol for *E. coli* Trx.

A quartz nanogravimetric method was implemented with the aim of calculating the water desorption as a function of temperature in a protein molecule. Figure 7 (panel A) reports the relative water desorption in the range from 25 to 145 °C for BacTrx and *E. coli* Trx; panel B shows the water desorption rates for both proteins, as obtained by computing the first derivative from the solid lines. The results demonstrated that BacTrx had a lower amount of bound water with respect to *E. coli* Trx, and that the temperature at which the maximum water desorption takes place was approx. 10 °C higher for BacTrx than for *E. coli* Trx.

Molecular modelling

The three-dimensional model of the oxidized form of BacTrx was constructed by a comparative molecular modelling technique based on satisfaction of spatial constraints, using *E. coli* Trx and *Anabaena* strain 7120 Trx as reference proteins. In a program-assisted analysis of secondary structural elements, no meaningful difference among the proteins was observed in the number of residues in β conformation, while the percentage of residues in helical conformation was 42% in BacTrx, 32% in *E. coli* Trx and 35% in *Anabaena* Trx. It is worth noting that the first helix

in BacTrx is longer than that of *E. coli* and *Anabaena* proteins as a consequence of the replacement of residues in loop conformation by residues involved in the helix structure; this implies the disappearance of a mobile region (residues 9–11 and 16–18 in *E. coli* Trx and 16–18 in *Anabaena* Trx). The decreased size of loop regions in BacTrx, to the advantage of the helical conformation, can be seen in other portions of the molecule, thus leading to an improved packing with minimization of thermal motion. Moreover, shorter loop regions are characterized by the presence of additional proline residues (Pro-19, Pro-83 and Pro-92), thus reducing the entropy of the unfolded state.

Replacements at positions 11, 12, 15, 91, 99 and 105 (numbering of BacTrx) by Gln should improve the packing in these regions of the structure because of hydrogen bonds (main chain–main chain Gln-11–Ile-14, Gln-11–Gln-15, Gln-12–Gly-16, Ala-101–Gln-105; side chain–side chain Gln-91–Gln-99).

The unfavourable helix–dipole interactions in the first and last turns of the helices, in *E. coli* and *Anabaena* Trxs, have been neutralized in BacTrx by appropriate replacements of given residues (Asp-15 and Glu-48 in *E. coli* protein and Glu-15 in *Anabaena* protein, at the C-terminus, and Lys-98 in *Anabaena* at the N-terminus of a helix are replaced in BacTrx by Gln, Ala and Gln residues respectively).

We found that the structures compared differ strongly in the number and the distribution of ion-pair interactions. There are ten ion-pairs in the structure of BacTrx, six in *E. coli* Trx and five in *Anabaena* Trx, involving 14/22 (64%), 7/27 (26%) and 10/27 (37%) of the charged residues respectively. The average number of ion-pairs per residue in proteins in general is 0.04 [28]; the number of ion-pairs per residue is 0.1 in BacTrx, 0.055 in *E. coli* Trx and 0.046 in *Anabaena* Trx. The extra salt bridges observed in the BacTrx model result from a change in sequence (relative to *E. coli* Trx and *Anabaena* Trx), and not from a change in local conformation. Figure 8 shows the structures presented as α -traces for BacTrx (A), *E. coli* Trx (B) and *Anabaena* Trx (C), with the relative ion-pair pattern indicated. In BacTrx the extra electrostatic interactions link secondary structural elements that are distant in the primary sequence. It is worth noting that the C-terminal helix is clamped by three salt bridges (Glu-41–Lys-93, Asp-102–Lys-85, Glu-97–His-46) to the rest of the BacTrx core. Furthermore, ion-pair bridges involving Asp-17, Asp-48 and Lys-18 in BacTrx show bidentate interactions; these kinds of bridges are not present in template proteins.

CONCLUSIONS

This paper reports the physicochemical features of the Trx of *Bacillus acidocaldarius*. The high-level expression system in *E. coli* is found to give active protein with a yield that averages 5% of total protein amount. Undoubtedly, the availability of pure protein in sufficient amounts facilitates certain investigations; indeed, NMR analyses of BacTrx are in progress to obtain information on its structure in solution.

Interestingly, BacTrx is the first Trx described so far from a thermophilic source. The results from CD and DSC measurements clearly demonstrated that BacTrx is endowed with a higher conformational heat stability when compared with the protein from *E. coli*. This evidence raises questions as to the molecular determinant(s) of this effect (see [32] for a review on molecular mechanism implicated in the thermal stability of proteins, and references therein).

In the sequence comparison with all bacterial Trxs, BacTrx does not emerge as unique for any of the amino acid ‘traffic rules’ for high-temperature molecular adaptation. Nanogravimetry revealed a minor content of bound water in BacTrx with

Figure 8 Ion-pairs patterns of BacTrx (A), *E. coli* Trx (B) and *Anabaena* strain 7120 Trx (C)

The structures in (B) and (C) are from [6] and [22] respectively. The structures are presented as C α traces. The side chains of positive, negative and active-site residues are pictured in blue, red and yellow respectively. Numbering is relative to each protein. (A) K18 N ϵ (NZ)...D17 O δ (OD)2, K79 NZ...D17 OD1, K54 NZ...D23 OD1, K93 NZ...E41 O ϵ (OE)2, K49 NZ...D48 OD1, K49 NZ...D48 OD2, K18 NZ...D48 OD2, K93 NZ...E94 OE2, H46 H ϵ (HE)2...E97 OE1, K85 NZ...D102 OD1. (B) K3 NZ...D2 OD2, K3 NZ...D47 OD1, K18 NZ...D13 OD2, K96 NZ...E44 OE1, K96 NZ...E48 OE1, K100 NZ...E48 OE2. (C) K18 NZ...E13 OE2, K54 NZ...E15 OE1, K69 NZ...E84 OE1, R79 NZ1...E91 OE2, H106 N ζ 2...D88 OD2.

respect to *E. coli* Trx, and a transition temperature approx. 10 °C higher for BacTrx. These results fit the recent concept that the reduction of the solvent-accessible surface is a structural feature of thermostable proteins [33]. Analysis of the three-dimensional model constructed for BacTrx showed a shortening of loop regions – with the eventual introduction of proline residues – and an increase in the number of ion-pairs, as significant differences in the comparison with the structures of *E. coli* and *Anabaena* Trxs. In some thermostable enzymes, reduced loops with ‘rigidifying’ proline residues have been observed that were absent in the mesophilic counterparts, and different studies comparing thermophilic and mesophilic proteins indicate an increase in the number of ion-pairs as a stabilizing factor [32].

The general belief that protein stability is due to many small effects, and that different factors stabilize different proteins, can be applied to the present investigation as well. An understanding of the molecular motifs of structural resistance in BacTrx could benefit from site-directed mutagenesis experiments; the expression system described here is highly suitable for the construction of mutant BacTrx molecules ‘tailored’ to perform structural and functional studies.

Professor Mario De Rosa (Istituto di Biochimica delle Macromolecole, Seconda Università di Napoli) is acknowledged for providing the bacterial paste of *B.*

acidocaldarius. We thank V. Carratore, F. Nozza and C. Rando for expert technical support. This work has been carried out under a research contract with Pharmacia-Upjohn (Milano, Italy), within the Programma Nazionale di Ricerca sulle Tecnologie per la Bioelettronica del Ministero dell’Università e della Ricerca Scientifica e Tecnologica.

REFERENCES

- Eklund, H., Gleason, F. K. and Holmgren, A. (1991) *Proteins* **11**, 13–28
- Loferer, H. and Hennecke, H. (1994) *Trends Biochem. Sci.* **19**, 169–171
- Holmgren, A. and Bjornstedt, M. (1995) *Methods Enzymol.* **252**, 199–208
- Buchanan, B. B., Schurmann, P., Decottignies, P. and Lozano, R. M. (1994) *Arch. Biochem. Biophys.* **314**, 257–260
- Jeng, M., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E. and Dyson, H. J. (1994) *Curr. Biol.* **2**, 853–868
- Kattii, S. K., LeMaster, D. M. and Eklund, H. (1990) *J. Mol. Biol.* **212**, 167–184
- LaVallie, E. R., Di Blasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. (1993) *Biotechnology* **11**, 187–193
- Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) *J. Biol. Chem.* **270**, 25328–25331
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) *Anal. Biochem.* **150**, 76–85
- Laemmli, U.K. (1970) *Nature (London)* **227**, 680–685
- Righetti, P. E. and Drysdale, J. W. (1976) in *Laboratory Techniques in Biochemistry and Molecular Biology (Work, T. S. and Work, E., eds.)*, vol. 5, pp. 562–565, North Holland Publishing, Amsterdam

- 12 Rabilloud, T., Vaillard, L., Gilly, C. and Lawrence, J. J. (1994) *Cell. Mol. Biol.* **40**, 57–75
- 13 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- 14 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- 15 Holmgren, A. (1979) *J. Biol. Chem.* **254**, 9627–9632
- 16 Barker, D. G. (1982) *Eur. J. Biochem.* **125**, 357–360
- 17 Saiki, R. K. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. A., Sninski, J. J. and White, T. J., eds.), pp. 13–20, Academic Press, New York
- 18 Sauerbrey, G. Z. (1964) *Z. Phys.* **178**, 457–462
- 19 Gascoyne, P. R. and Pething, R. (1977) *J. Chem. Soc., Faraday Trans.* **73**, 171–180
- 20 Nicastro, G., de Chiara, C., Tato, M., Franzoni, L., Carrara, S., Carrara, E. and Nicolini, C. (1995) *Q. Magn. Res. Biol. Med.* **4**, 283–298
- 21 Erokhin, V., Facci, P. and Nicolini, C. (1995) *Biosensors Bioelectron.* **10**, 25–34
- 22 Saarinen, M., Gleason, F. K. and Eklund, H. (1995) *Structure* **3**, 1097–1108
- 23 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542
- 24 Sander, C. and Schneider, R. (1991) *Proteins* **9**, 56–68
- 25 Sali, A. and Blundell, T. J. (1993) *J. Mol. Biol.* **234**, 779–815
- 26 Laskowski, R. A., MacArthur, M. W., Moss, M. W. and Thornton, J. M. (1993) *J. Appl. Cryst.* **26**, 283–291
- 27 Vriend, G. (1990) *J. Mol. Graphics* **8**, 52–56
- 28 Barlow, D. J. and Thornton, J. M. (1983) *J. Mol. Biol.* **168**, 857–885
- 29 Kabsch, W. and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- 30 Holmgren, A. (1972) *J. Biol. Chem.* **247**, 1992–1998
- 31 Ladbury, J. E., Wynn, R., Hellinga, H. W. and Sturtevant, M. (1993) *Biochemistry* **32**, 7526–7530
- 32 Vielle, C., Burdette, D. S. and Zeikus, J. G. (1996) *Biotechnol. Annu. Rev.* **2**, 1–83
- 33 Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. and Rice, D. W. (1995) *Structure* **3**, 1147–1158
- 34 Hoog, J.-O., vonBahr-Lindstrom, H., Josephson, S., Wallace, B., Kushner, S. R., Jornvall, H. and Holmgren, A. (1985) *Biosci. Rep.* **4**, 917–923
- 35 McFarlan, S. C., Hogenkamp, H. P. C., Eccleston, E. D., Howard, J. B. and Fuchs, J. A. (1989) *Eur. J. Biochem.* **179**, 389–398
- 36 Meng, M. and Hogenkamp, H. P. C. (1981) *J. Biol. Chem.* **256**, 9174–9182
- 37 Gleason, F. K., Whittaker, M. M., Holmgren, A. and Jornvall, H. (1985) *J. Biol. Chem.* **260**, 9567–9573
- 38 Alam, J., Curtis, S., Gleason, F. K., Gerami-Nejad, M. and Fuchs, J. A. (1989) *J. Bacteriol.* **171**, 162–171
- 39 Muller, E. G. D. and Buchanan, B. B. (1989) *J. Biol. Chem.* **264**, 4008–4014
- 40 Mathews, W. R., Johnson, R. S., Cornwell, K. L., Johnson, T. C., Buchanan, B. B. and Biemann, K. (1987) *J. Biol. Chem.* **262**, 7537–7545
- 41 Johnson, R. S. and Biemann, K. (1987) *Biochemistry* **26**, 1209–1214
- 42 Clement-Metral, J., Holmgren, A., Cambillau, C., Jornvall, H., Eklund, H., Thomas, D. and Lederer, F. (1988) *Eur. J. Biochem.* **172**, 413–419
- 43 Johnson, T. C., Yee, B. C., Carlson, D. E., Buchanan, B. B., Johnson, R. S., Mathews, W. R. and Biemann, K. (1988) *J. Bacteriol.* **170**, 2406–2408