Extremophiles (1998) 2:61–66

ORIGINAL PAPER

Helen Connaris · Shauna M. West · David W. Hough
Michael J. Danson

Cloning and overexpression in *Escherichia coli* of the gene encoding citrate synthase from the hyperthermophilic Archaeon *Sulfolobus solfataricus*

Received: August 10, 1997 / Accepted: October 23, 1997

Abstract The citrate synthase (CS) gene from the hyperthermophilic Archaeon *Sulfolobus solfataricus* has been cloned and sequenced. The gene encodes a polypeptide of 378 amino acids with a calculated polypeptide molecular mass of 42679. High-level expression was achieved in *Escherichia coli* and the recombinant citrate synthase was purified to homogeneity using a heat step and dye-ligand affinity chromatography. This procedure yielded approximately 26 mg of pure CS per liter of culture, with a specific activity of 41 U/mg. The enzyme exhibited a half-life of 8 min at 95°C. A homology-modelled structure of the *S. solfataricus* CS has been generated using the crystal structure of the enzyme from the thermoacidophilic Archaeon *Thermoplasma acidophilum* with which it displays 58% sequence identity. The modelled structure is discussed with respect to the thermostability properties of the enzyme.

Key words Thermophile · Citrate synthase · Thermostability · Archaea · Gene sequence · *Sulfolobus*

Introduction

Citrate synthase (EC 4.1.3.7) catalyzes the condensation reaction of oxaloacetate and acetyl-CoA to form citrate and CoA, thus effecting the entry of carbon into the citric acid cycle. The enzyme occurs throughout the domains Eukarya, Bacteria, and Archaea (Woese et al. 1990), and we have chosen it as a model system with which to investigate the molecular basis of protein stability in extreme environments.

We have previously cloned, sequenced, and expressed in *Escherichia coli* the citrate synthase (CS) gene from the thermophilic Archaea *Thermoplasma acidophilum* and *Pyrococcus furiosus*, which grow optimally at 55°C and 100°C, respectively (Sutherland et al. 1990, 1991; Muir et al. 1995). Crystal structures have been determined for these dimeric enzymes and comparisons have been made with respect to the mesophilic dimeric CS from pig (Remington et al. 1982) to identify possible structural features that confer stability to high temperatures (Russell et al. 1994, 1997). The archaeal CSs possess more compact subunits than the mesophilic enzyme, this being achieved by shortening of loops, an increase in the number of atoms buried, optimized packing of side-chains in the interior of the protein, and a reduction in the number and size of internal cavities. Also, they exhibit a more intimate association of their subunits, through a greater complementarity between the monomers and, in the *Pyrococcus* enzyme, by the C-terminal region of each monomer folding over the surface of the other monomer. In the hyperthermophilic CS from *Pyrococcus*, there is an increase in complex ion-pair networks, especially at the subunit interface, whereas the thermophilic enzyme from *Thermoplasma* has a markedly increased hydrophobicity in the intersubunit contact area. Finally, a reduction in the content of the thermolabile residues Asn, Gln, and Cys, and of Asp, which can undergo chain-cleavage reactions at elevated temperatures, is found in the thermophilic CSs.

Sulfolobus solfataricus, a hyperthermophilic Archaeon that grows at 80°C, also possesses a dimeric CS (Lill et al. 1992). In the present paper, we report the cloning, sequencing, and expression of the CS gene from *S. solfataricus*. The characterization of the recombinant enzyme is reported and a homology-modelled structure is generated and analyzed.

Materials and methods

Strains and plasmid vectors

Cells of *Sulfolobus solfataricus* (strain DSM 1616) were supplied by Dr. N. Raven (Centre for Applied Microbiology and Research, Porton Down, UK). *E. coli* strain JM109

Communicated by G. Antranikian

H. Connaris (✉) · S.M. West · D.W. Hough · M.J. Danson
Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, UK
Tel. +44-1225-826509; Fax +44-1225-826779
e-mail: H.Connaris@bath.ac.uk

and *E. coli* XL-1 MRA Blue (Promega, Southampton, UK) were used for DNA cloning and gene expression. The cloning vector pGEM-T was obtained from Promega. The expression vector pREC7*Nde*I was kindly donated by Dr. L. Kurz (University of Ohio, USA).

Amplification of DNA by the polymerase chain reaction (PCR)

A partial coding sequence of the CS gene from *S. solfataricus* was obtained from genomic DNA by PCR amplification. Using the known codon usage for the *Sulfolobus* genome (Zillig et al. 1980), degenerate primers were designed from the N-terminal amino acid sequence (positions 9–16) of the largest tryptic peptide fragment yielded from limited proteolysis of pure *S. solfataricus* CS (Lill et al. 1992), and from a conserved region (HRVYKTYD) of all known CS sequences. Amplification was carried out using 500 ng of *S. solfataricus* genomic DNA, prepared as described by Yeats et al. (1982), 1 μM of each primer, and 2.5 U of Biotaq polymerase (Bioline, London, UK). The initial cycle for denaturation was at 94°C for 4 min, followed by 30 cycles of 94°C for 90 s, 46°C for 90 s, and 72°C for 4 min. A further extension cycle of 7 min was performed. The reaction products were separated by electrophoresis on a 0.7% (w/v) agarose gel and the expected 800-bp fragment was purified using Gene Clean II kit (BIO 101, Vista, CA, USA), cloned into pGEM-T, and sequenced to confirm its identity.

Construction of a genomic DNA library

A genomic DNA library for *S. solfataricus* was constructed in λEMBL3 (Stratagene, Cambridge, UK). Partially digested chromosomal DNA (200 μg), after *Sau*3A1 treatment at 37°C for 30 min, was fractionated on a 5%–25% (w/v) NaCl gradient, and inserts of approximately 15 kb were pooled and ligated to 1 μg of *Bam*HI arms of λEMBL3. The mixture was packaged according to the manufacturer's instructions, and *E. coli* XL1-Blue MRA(P2) was infected with the packaging mix. Screening was carried out as described by Sambrook et al. (1989) using the 800-bp PCR gene product randomly labelled with ^{32}P. Positive plaques were isolated and used in a secondary screen, before λDNA was prepared according to the manufacturer's instructions (Qiagen, Crawley, UK). Direct sequencing of λDNA was performed on a Perkin Elmer ABI Prism 377 DNA sequencer using gene-specific primers.

Heterologous gene expression and purification of the recombinant citrate synthase

The expression of the CS gene from *S. solfataricus* was performed using the vector pREC7*Nde*I, which relies on the *rec*A promoter for high-level expression induced with nalidixic acid (Ogawa et al. 1979). The CS gene was amplified from a λ clone using gene-specific primers based on the 5′ and 3′ ends of the coding region and incorporating restriction sites (in bold type) to facilitate the cloning of the gene into the expression vector. The primers employed were A2a(PCR1), 5′GACTGA**CATATG**AGTGTCGTA3′, which introduces an *Nde*I site and conserves the initiation codon, and primer A2a(PCR2), 5′CTATTAAAG**GCATGC**ATGTAT3′, which introduces a *Sph*I site. Amplification was carried out as before, but this time using 100 ng of λA2a template containing the *S. solfataricus* CS gene, and annealing at 55°C instead of 46°C. The 1.13-kb gene product was extracted from 0.8% (w/v) agarose gels as before, digested with *Nde*I and *Hind*III, and ligated into pREC7*Nde*I, before transforming *E. coli* JM109. Cultures (1 liter) were grown in Terrific Broth (Sambrook et al. 1989) containing ampicillin (100 μg/ml) to an $OD_{600\,nm}$ of 1.1, before induction with nalidixic acid (50 μg/ml) for 18–20 h. Cells were harvested, sonicated, and heat-treated for 20 min at 75°C to precipitate host cell proteins. CS was purified using a Matrex Gel Red A column, as described by James et al. (1994). Enzyme purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a 10% (w/v) polyacrylamide gel (Laemmli 1970).

Assay of citrate synthase

CS activity was assayed spectrophotometrically at 412 nm by the method of Srere et al. (1963). The assay was performed at 55°C in 20 mM Tris-HCl buffer (pH 8.0), containing 2 mM ethylenediaminetetraacetic acid (EDTA), 0.15 mM acetyl-CoA, 0.2 mM oxaloacetate, and 0.1 mM 5,5′-dithiobis(2-nitrobenzoic acid). One unit of enzyme is defined as one μmole of CoA produced per minute. For the determination of K_m and V_{max} values, initial rate measurements were performed at varying substrate concentrations of acetyl-CoA and oxaloacetic acid, and the data were analyzed by the direct linear plot of Eisenthal and Cornish-Bowden (1974).

Thermostability analysis

The thermal stability of native and recombinant *Sulfolobus* CS was determined by pre-incubating pure samples in sealed glass vials at temperatures up to 100°C. Aliquots were taken at various time intervals, cooled rapidly on ice, and assayed for residual CS activity under standard conditions (Srere et al. 1963).

Sequence alignment and molecular modelling

A structurally based sequence alignment was constructed with the CS sequences from *S. solfataricus*, *P. furiosus, T. acidophilum*, and pig, using AMPS (Barton and Sternberg 1990) and ALSCRIPT (Barton 1993); the crystal structures of the CSs from the latter three organisms were used to define regions of secondary structure. The primary sequence of *S. solfataricus* CS was then homology-modelled to the crystal structure of the open (non ligand-bound) form

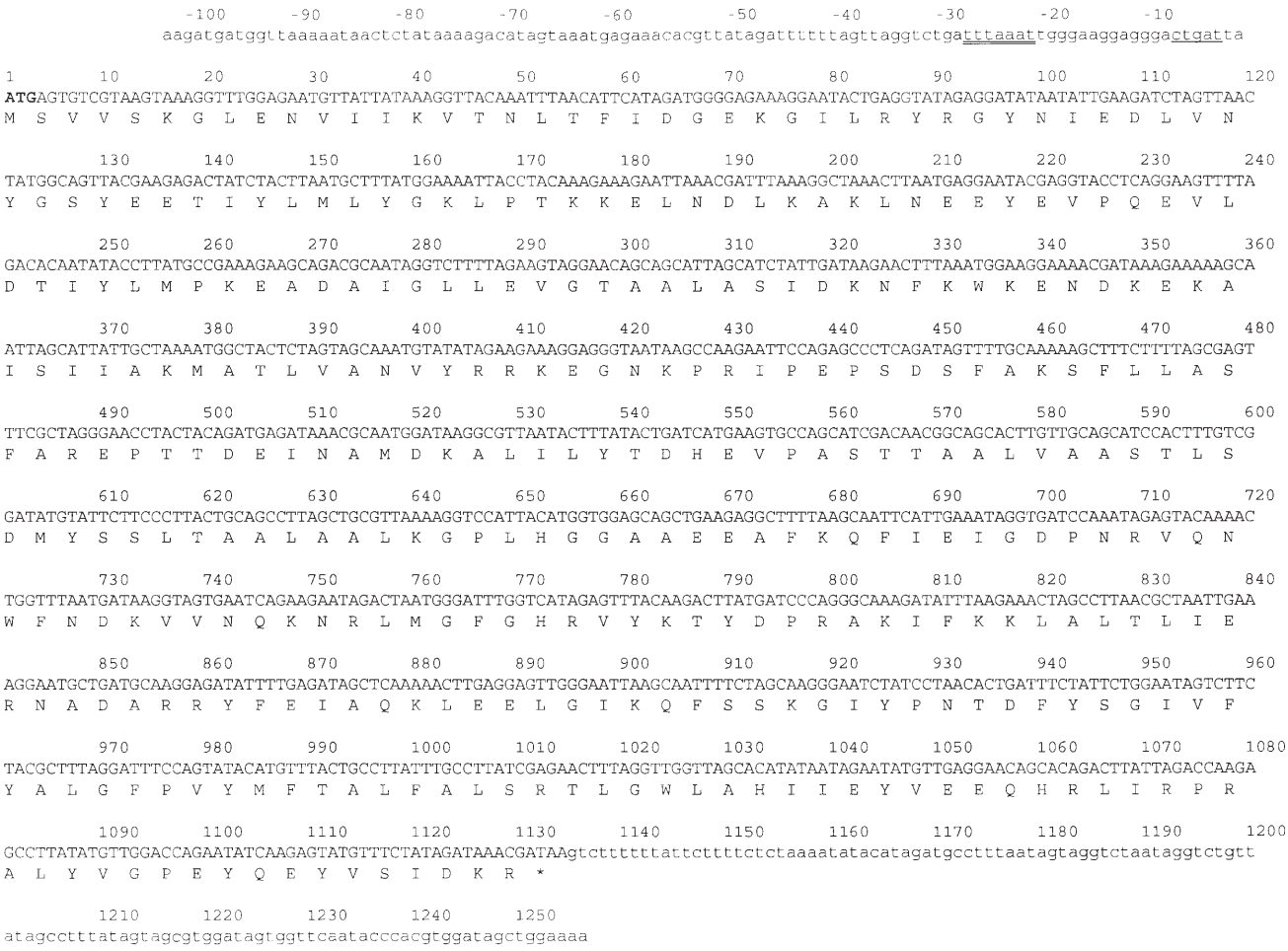

Fig. 1. Nucleotide sequence and deduced amino acid sequence of the gene encoding citrate synthase (CS) from *Sulfolobus solfataricus*. The open reading frame is indicated at *nt* positions 1 to 1131. The location of a putative ribosome binding site is *underlined* at *nt* positions −6 to −2, and an archaeal proximal consensus sequence is doubly underlined at *nt* positions −26 to −20

of *T. acidophilum* CS at 0.25 nm using O version 5.0 (Jones et al. 1991). Equivalent residues were identified using the structurally based sequence alignment, where all deletions/insertions were localized in loop regions. All atoms with unknown coordinates were built with ideal geometries. Only a single monomer was modelled, with the dimer generated by applying a two-fold rotation to the monomer. The resulting model was assessed by ENVIRONMENTS (Luthy et al. 1992).

Nucleotide sequence accession number

The sequence reported in this study has been deposited in the GenBank database under the accession number U70879.

Results and discussion

Cloning and sequencing of *S. solfataricus* CS gene

The partial coding sequence of the *S. solfataricus* CS gene was amplified by PCR. This yielded a PCR product of approximately 800 bp, which was subsequently sequenced and, based on alignments with other known citrate synthases, was confirmed to be a partial gene encoding citrate synthase. The 800-bp DNA fragment was radiolabelled with ^{32}P and used as a probe to screen a λEMBL3 library constructed from partially digested *S. solfataricus* genomic DNA. The probe was hybridized under stringent conditions (hybridization temperature 60°C) and 8 positive recombinants were identified in a screen of 400 plaques. Using gene-specific primers and direct sequencing of lambda DNA from these plaques, one plaque, designated as λA2a, was found to contain the complete *S. solfataricus* CS gene on an 11-kb fragment, with sequence confirmed on both strands.

The sequence of a 1353-bp region of *S. solfataricus* genomic DNA comprising the complete CS gene and its deduced amino acid sequence is shown in Fig. 1. The gene encodes a polypeptide of 378 amino acids with an M_r of 42 679, which is similar in size to the protein purified by Lill et al. (1992). The calculated G + C content of the coding sequence was 38%, reflecting the base composition (34% G + C) of the *Sulfolobus* genome (Zillig et al. 1980). A putative ribosome binding site (CTGAT, position −6 to

Table 1. Purification of the recombinant *Sulfolobus solfataricus* citrate synthase from *Escherichia coli* JM109

Stage	Volume (ml)	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (fold)
Cell extract	7	2380	325	7	100	–
Heat treatment	5	1800	63	29	76	4
Red Gel A	6	1070	26	41	45	6

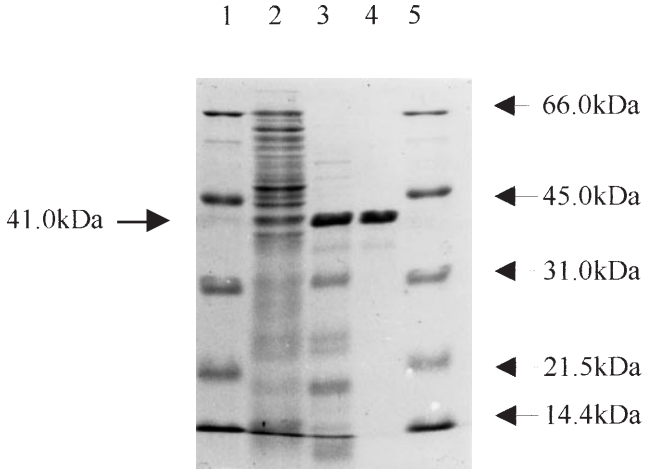

Fig. 2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) analysis of recombinant *S. solfataricus* CS. The samples are recombinant *Escherichia coli* cell extract (*lane 2*), cell extract after heat treatment (*lane 3*), pooled fractions from Matrex Red Gel A column (*lane 4*), and standard protein markers (*lanes 1 and 5*)

−2) and archaeal proximal promoter (TTTAAAT, position −26 to −20) could be identified upstream from the initiation site. An archaeal transcription termination region has been located downstream of the stop codon (TAA, position 1129), consisting of long stretches of pyrimidine-rich sequences. Surprisingly, the amino acid sequence of the N-terminal 30 residues was only 67% identical to that determined from the purified enzyme by Lill et al. (1992). We currently have no explanation for this discrepancy.

Expression and purification of recombinant CS

High-level expression was achieved when cultures of *E. coli* JM109, transformed with the expression vector pREC7*Nde*I containing the CS coding region from *S. solfataricus*, were induced with nalidixic acid. The enzyme was expressed as a soluble active recombinant CS, which was purified in two stages to yield approximately 26 mg of protein per liter of culture, with a specific activity of 41 U/mg and an M_r of 41,000 (Table 1, Fig. 2). The native and recombinant enzymes were comparable with respect to their specific activities (33 ± 2 and 41 ± 8 U/mg) and their K_m values for acetyl CoA (2.1 ± 0.1 μM and 2.3 ± 0.2 μM) and oxaloacetate (4.7 ± 1.2 μM and 5.1 ± 1.4 μM). Also, both enzymes were equally resistant to thermal inactivation, with half-lives of 8 min at 95°C.

Sequence alignment and homology modelling of *S. solfataricus* CS

A structurally based sequence alignment of the *S. solfataricus* CS with the CSs from *T. acidophilum*, *P. furiosus*, and pig is shown in Fig. 3. *Sulfolobus* CS shows close similarity to the CSs from the thermophilic Archaea *T. acidophilum* and *P. furiosus* with 58% and 45% identity, respectively, while its identity with pig CS is substantially lower (34% identity). As noted previously for other archaeal CSs (Muir et al. 1994), the *Sulfolobus* enzyme is 37 residues shorter at the N-terminus than the pig protein. However, of the 11 residues implicated in the catalytic action of the pig enzyme, 8 are conserved in the archaeal CSs. Using this information, and its high homology with *T. acidophilum* CS, the CS from *S. solfataricus* was homology-modelled using the crystal structure of *T. acidophilum* resolved to 0.25 nm as a template (Russell et al. 1994). In turn, the model was superimposed with the CSs from *P. furiosus* (Russell et al. 1997) and pig (Remington et al. 1982) to determine any structural differences between the different enzymes.

All four CSs are structurally homologous despite their limited sequence identity, with similar active-site structures and overall folds. In comparison with the pig enzyme, the three thermophilic archaeal CSs show a marked reduction in the length of loops, and a shortening of the N-terminal arm (helix A), both of which lead to more compact structures. Other features that might lead to even more compact proteins, such as better residue packaging and a reduction in the size and number of cavities, cannot be determined from the modelled structure of the *Sulfolobus* CS, and analyses of these await crystal structure data.

The intersubunit interactions in the *Sulfolobus* CS can be assessed with confidence from the modelled structure. In the pig, *T. acidophilum* and *P. furiosus* CS crystal structures, the intersubunit interface is composed of helices F,G,L, and M arranged as four antiparallel pairs. In the CS from pig, helices F, L, and M are hydrophobic on the sides facing the other monomer, whereas helix G is hydrophilic. In the *T. acidophilum* CS, all four helices are hydrophobic, there being a significant increase in the alanine residues in this region of the thermostable enzyme that we have sug-

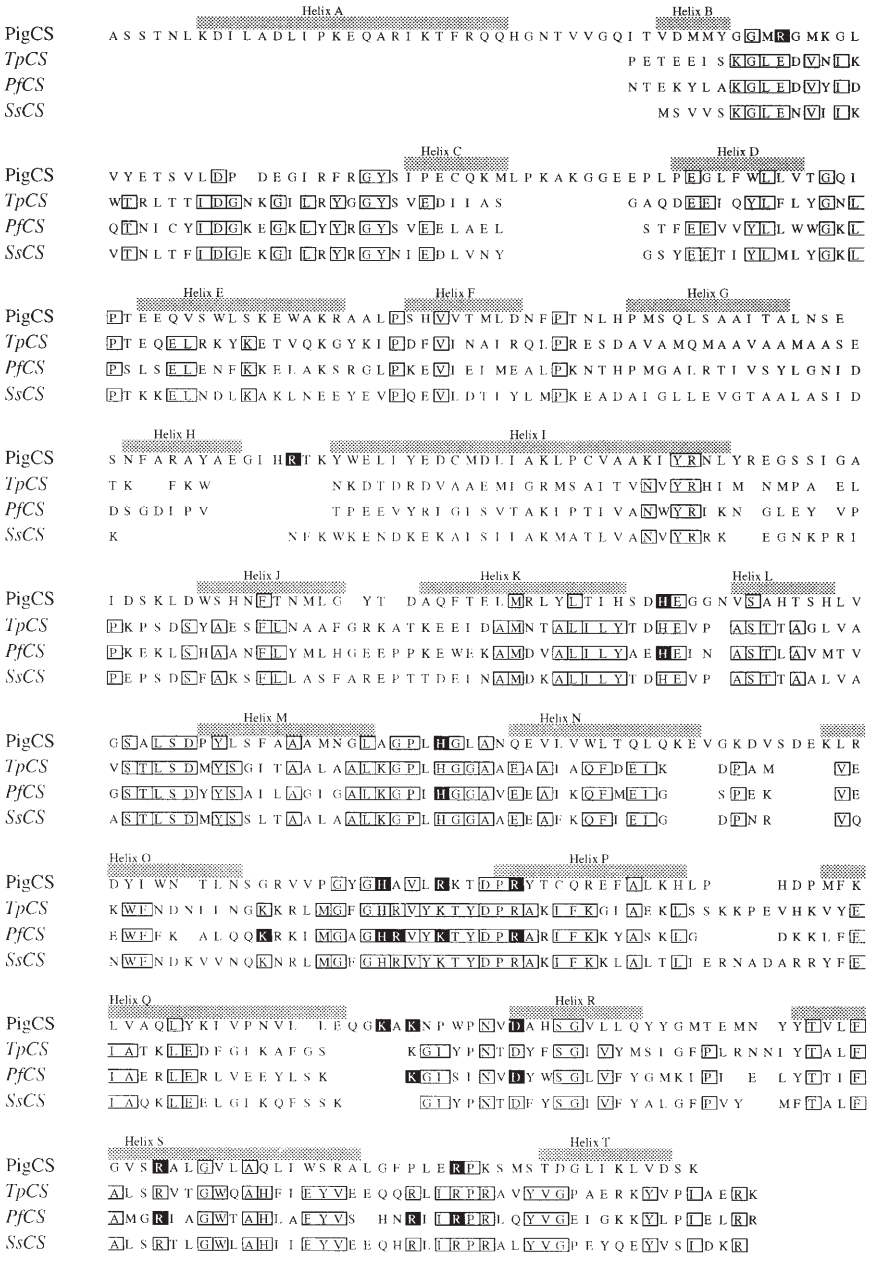

Fig. 3. Structurally based sequence alignment of the translated amino acid sequence of *S. solfataricus* CS (SsCS) with those from *P. furiosus (PfCS)*, *T. acidophilum (TpCS)*, and pig (*PigCS*). The *bar above* the sequences indicates regions of secondary structure. Residues implicated in the catalytic and binding sites from defined crystal structures of pig and *P. furiosus* CS are *highlighted in black*. Identities across all three archaeal sequences are *boxed*

gested may contribute to maintaining structural integrity at elevated temperatures (Russell et al. 1994). Interestingly, at the subunit interface of the *P. furiosus* CS, we have located a five-membered ion-pair network at each end of helices G and M, and an additional four-membered intersubunit network outside the four-helix bundle (Russell et al. 1997). The presence of ion-pair networks is a feature common to the seven available crystal structures of enzymes from hyperthermophilic Archaea and is strongly implicated in their inherent stability.

The subunit interface of *S. solfataricus* CS also comprises helices F,G,L, and M arranged as four antiparallel pairs. As with the *T. acidophilum* CS, there is an increase in helix-stabilizing alanine residues in helix G, which would result in an increase in hydrophobic interactions when compared with the pig and *P. furiosus* enzyme structures. However, in helix G, Met 99 of *T. acidophilum* CS is a glutamate in the *S. solfataricus* enzyme and it is likely that this Glu may form an additional ion pair at the intersubunit interface. Therefore, in this respect, the *S. solfataricus* is intermediate between the *T. acidophilum* and *P. furiosus* enzymes, as indeed is its thermostability.

Finally, with respect to the content of thermolabile amino acids, the *S. solfataricus* CS contains no cysteine residues and there is a decrease of glutamines with increasing temperature across the CS family (pig, 17 residues; *T. acidophilum*, 10; *S. solfataricus*, 8; and *P. furiosus*, 5). However, as pointed out by Hensel (1993), hyperthermophilic proteins still contain significant numbers of the amino acids that undergo thermal degradation at temperatures >85°C,

and there is evidence that the folded conformation of the proteins protects these residues from destruction at physiological temperatures.

Concluding remarks

The work reported in this paper gives a family of citrate synthases from organisms growing optimally at 37°C, 55°C, 80°C, and 100°C. The data suggest that the *S. solfataricus* CS is structurally homologous to the other members of this family and that, with the eventual determination of its atomic structure, this system will allow the detailed exploration of the structural basis of protein thermostability and of the balance between the maintenance of structural integrity at elevated temperatures and the need for conformational flexibility for effective catalysis.

Acknowledgments We thank the Biotechnology and Biological Sciences Research Council (UK) for a Research Grant to MJD, DWH, and Dr. Garry Taylor. We are also indebted to Dr. Taylor and Dr. Rupert Russell for their help and advice, especially with the homology modelling.

References

Barton GJ (1993) ALSCRIPT – a tool to format multiple sequence alignments. Protein Eng 6:37–40

Barton GJ, Sternberg MJE (1990) Flexible protein sequence patterns: a sensitive method to detect weak structural similarities. J Mol Biol 212:389–402

Eisenthal R, Cornish-Bowden A (1974) The direct linear plot. Biochem J 139:715–721

Hensel R (1993) Proteins of extreme thermophiles. New Comp Biochemistry [The Biochemistry of Archaea] 26:209–221 (Eds. M. Kates, D. Kushner and A.T. Matheson) Elsevier/North Holland Press, Amsterdam

James KD, Russell RJM, Parker L, Daniel RM, Hough DW, Danson MJ (1994) Citrate synthases from the Archaea: Development of a bio-specific, affinity chromatography purification procedure. FEMS Microbiol Lett 119:181–186

Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building models in electron density maps and location of errors in these models. Acta Crystallog. A 47:110–119

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head, of bacteriophage T4. Nature 227:680–685

Lill U, Lefrank S, Henschen A, Eggerer H (1992) Conversion, by limited proteolysis, of an archaebacterial citrate synthase into essentially a citryl-CoA hydrolase. Eur J Biochem 208:459–466

Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with 3-dimensional profiles. Nature 356:83–85

Muir JM, Hough DW, Danson MJ (1994) Citrate synthases from Archaea. Syst Appl Microbiol 16:528–533

Muir JM, Russell RJM, Hough DW, Danson MJ (1995) Citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus*. Protein Eng 8:583–592

Ogawa T, Wabico H, Tsurimoto T, Horri T, Masukata H, Ogawa H (1979) Characteristics of purified RecA protein and the regulation of the synthesis in vivo. Quant Biol 43:909–915

Remington S, Wiegand G, Huber R (1982) Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. J Mol Biol 158:111–152

Russell RJM, Hough DW, Danson MJ (1994) The crystal structure of citrate synthase from the thermophilic Archaeon *Thermoplasma acidophilum*. Structure 2:1157–1167

Russell RJM, Hough DW, Danson MJ, Taylor GL (1997) The crystal structure of citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochemistry 36:9983–9994

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Srere PA, Brazil H, Gonen L (1963) The citrate condensing enzyme of pigeon breast muscle and moth flight muscle. Acta Chem Scand 17:S129–S134

Sutherland KJ, Henneke CM, Towner P, Hough DW, Danson MJ (1990) Citrate synthase from the thermophilic archaebacterium *Thermoplasma acidophilum*. Cloning and sequencing of the gene. Eur J Biochem 194:839–844

Sutherland KJ, Danson MJ, Hough DW, Towner P (1991) Expression and purification of plasmid-encoded *Thermoplasma acidophilum* citrate synthase from *Escherichia coli.* FEBS Lett 282:132–134

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Yeats S, McWilliam P, Zillig W (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO J 1:1035–1038

Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz J (1980) The *Sulfolobus*-"Caldariella" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch Microbiol 125:259–269