Acta Crystallographica Section F
**Structural Biology and Crystallization Communications**
ISSN 1744-3091

# Expression, purification, crystallization and preliminary X-ray characterization of two crystal forms of stationary-phase survival E protein from *Campylobacter jejuni*

**A. M. D. Gonçalves,[a] A. T. Rêgo,[a] M. Thomaz,[a,b] F. J. Enguita[a,c]* and M. A. Carrondo[a]**

[a]Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Avenida da República, Apartado 127, 2781-901 Oeiras, Portugal, [b]Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2748-901 Oeiras, Portugal, and [c]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

Correspondence e-mail: fenguita@fm.ul.pt

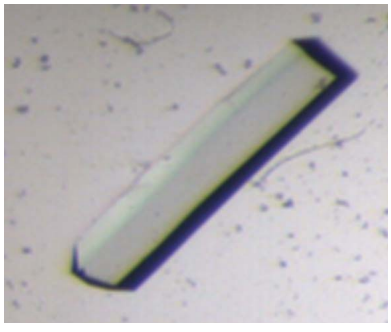
Received 8 October 2007
Accepted 28 January 2008

Survival E (SurE) protein from *Campylobacter jejuni*, a Gram-negative mesophile, has been overexpressed in *Escherichia coli* as a soluble protein, successfully purified and crystallized in two distinct crystal forms. The first form belongs to space group $P2_12_12_1$, with a tetramer in the asymmetric unit and unit-cell parameters $a = 80.5$, $b = 119.0$, $c = 135.3$ Å. The second form belongs to space group $C2$, with unit-cell parameters $a = 121.4$, $b = 47.1$, $c = 97.8$ Å, and contains a dimer in the asymmetric unit. Diffraction data have been collected from these crystal forms to 2.5 and 2.95 Å resolution, respectively.

## 1. Introduction

As growth-supporting substrates are depleted from their environment, bacterial cells enter a phase defined as the stationary phase of growth. In *Escherichia coli*, the stationary-phase alternative sigma factor controls the expression of genes involved in cell survival in response to cessation of growth and provides cross-protection against various stresses. This stationary-phase RNA polymerase subunit ($\sigma^s$) is encoded by the *rpo*S gene. In *E. coli* this gene is closely related to three other genes, all of which are involved in stress-response phenotypes (Loewen & Hengge-Aronis, 1994). These genes cluster into a stationary-phase stress-survival operon *surE-pcm-nlpD-rpoS*, in which the *nlpD* gene encodes an outer membrane lipoprotein, while the *pcm* gene product, protein-L-isoaspartate-*O*-methyltransferase, is involved in protein repair by conversion of deaminated or isomerized aspartyl residues to L-aspartyl residues (Li *et al.*, 1997; Visick *et al.*, 1998). The *surE* and *pcm* genes overlap by four base pairs, forming a bicistronic region of the operon. Their co-expression is believed to represent parallel responses *E. coli* has developed against protein ageing (Visick *et al.*, 1998); this is sustained by the observation that *surE*/*pcm* double mutants have a much greater deleterious effect on cell viability than either single-knockout mutant.

The precise physiological and biochemical function of the *surE* gene (gi 6967767) is unknown. This gene encodes a polypeptide of 258 amino acids with a calculated MW of 29 013 Da and a theoretical isoelectric point of 9.3. This evolutionarily conserved protein was initially proposed to function as an acid phosphatase. More recently, its enzymatic activity was shown to be towards various nucleotides (5′- or 3′-monophosphates, purine and pyrimidine ribonucleotides and deoxyribonucleotides) as well as inorganic phosphate, thus revealing broad substrate specificity; therefore, it was renamed and reclassified as a nucleotidase. The enzymatic function of SurE is also metal-dependent, showing various activities in the presence of divalent cations such as $Mg^{2+}$, $Mn^{2+}$, $Co^{2+}$, $Ca^{2+}$ and $Zn^{2+}$ (Lee *et al.*, 2001; Proudfoot *et al.*, 2004). In all SurE-like proteins annotated in sequenced genomes, the metal-binding site contains the highly conserved N-terminal NDDG motif, in which the two conserved aspartates, together with other strictly conserved serine and asparagine residues, coordinate the divalent cation with an octahedral geometry (Lee *et al.*, 2001; Zhang *et al.*, 2001; Mura *et al.*, 2003).



Despite thorough enzymatic characterization of the *E. coli* and archaeal SurE proteins, they generally displayed rather low phosphatase and nucleotidase activities towards substrates in all studies; hence, their true biological substrate still remains unknown. The availability of an ever-growing number of sequenced microbial genomes has allowed the identification of new classes of enzymes that are active towards noncanonical nucleotides, some of which constitute toxic byproducts of cellular metabolism arising from oxidation, inappropriate methylation or deamination (Kamiya, 2003) of DNA. Such enzymes perform a 'housekeeping' function by maintaining the cellular chemical pool free of abnormal metabolites that would otherwise produce mispairing nascent DNA and severely affect the mutation rate (Galperin *et al.*, 2006). In this respect, the SurE enzyme may represent a possible candidate for a 'housekeeping' enzyme.

Here, we report the cloning, expression, purification, crystallization and preliminary crystallographic data of *Campylobacter jejuni* stationary-phase survival E protein (*Cj*SurE). This enzyme shows a relatively high homology to the SurE enzymes from *Thermatoga maritima* (*Tm*SurE; 36%) and *Pyrobaculum aerophilum* (34%). Despite the sequence similarity, *Cj*SurE presents a notably high isoelectric point (9.8), in contrast to the acidic values for the *T. maritima* (4.77) and *P. aerophilum* (6.15) enzymes. Although the *E. coli* homologue was the first to be genetically and biochemically characterized, the only structures available are those from the extremophiles *T. maritima* and *P. aerophilum*, an archaeabacterium (Lee *et al.*, 2001; Zhang *et al.*, 2001; Mura *et al.*, 2003). More recently, the crystal structure of SurE from *Thermus thermophilus* has been reported (Iwasaki & Miki, 2007). Therefore, *C. jejuni* stationary-phase survival E protein will be the first mesophilic protein from this family to be characterized by X-ray crystallography.

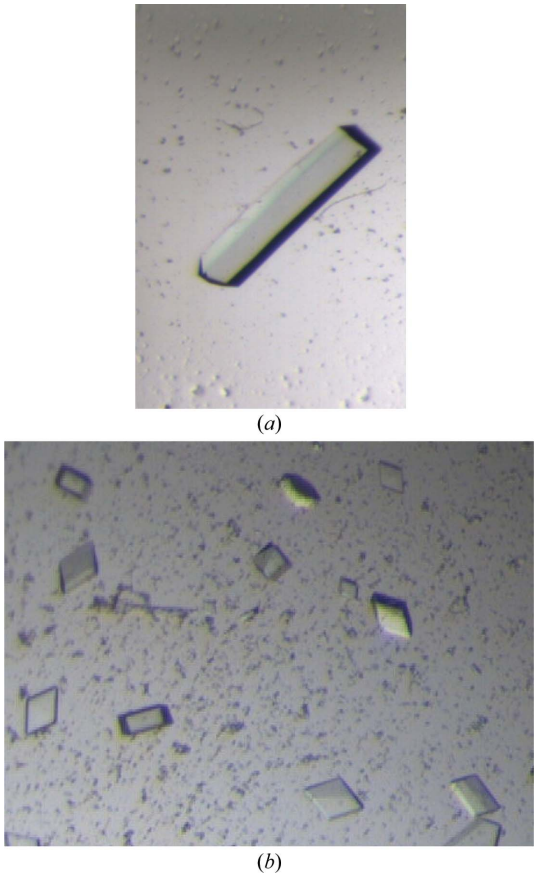


**Figure 1**
Crystals of *C. jejuni* stationary-phase survival E protein grown in (*a*) 0.02 *M* sodium citrate, 0.1 *M* sodium dihydrogen phosphate pH 6.2 and 15%(*w*/*v*) PEG 2000 and (*b*) 0.2 *M* magnesium acetate and 20%(*w*/*v*) PEG 3350. The longest dimension of the crystal in (*a*) is 0.2 mm, while the dimensions of the crystals in (*b*) were typically 0.05 × 0.06 mm.

## 2. Experimental

### 2.1. Cloning, expression and purification of *Cj*SurE

Genomic DNA from *C. jejuni* NCTC 11168 strain was used as a template from which the *surE* gene (gi 6967767) was isolated using the polymerase chain reaction (PCR). Oligonucleotide primers were designed to introduce *Nco*I and *Xho*I restriction sites to aid in cloning with the initiation sequence (ATG) within the site. The full-length *surE* gene (corresponding to residues 1–258) was cloned into pETM-10 vector (a kind gift from the Protein Expression and Purification Facility, EMBL Heidelberg, Germany). Recombinant *Cj*SurE was expressed as a fusion with an N-terminal hexahistidine tag (MKHHHHHHP).

The expression constructs were transformed into the *E. coli* BL21 (DE3) host strain (Stratagene) using a heat-shock technique. Expression was carried out in 1–2 l batches of Luria–Bertani enriched growth medium containing 50 µg ml$^{-1}$ kanamycin, which were inoculated with a sample of an overnight culture. Cultures were grown at 310 K until an optical density of approximately 0.6–0.7 at 600 nm was reached, at which point IPTG (isopropyl $\beta$-D-1-thiogalactopyranoside) was added to a final concentration of 0.3 m*M*, thus inducing overexpression. Cultures were allowed to grow for a further 5 h, after which cells were harvested at 7000*g* for 20 min at 277 K and resuspended in 10 ml lysis buffer [25 m*M* Tris–HCl pH 7.0, 500 m*M* NaCl, 7 m*M* PMSF, 2 m*M* $\beta$-mercaptoethanol, 5%(*v*/*v*) glycerol] per litre of culture. Cells were frozen at 193 K, thawed and lysed by two passages through a French Press. DNase I was added to a final concentration of 20 µg ml$^{-1}$. The lysed culture was centrifuged at 40 000*g* for 30 min at 277 K.

The soluble fraction was then loaded onto a 5 ml HiTrap $Ni^{2+}$-affinity column (Amersham-Pharmacia) previously equilibrated with loading buffer [25 m*M* Tris–HCl pH 7.0, 500 m*M* NaCl, 2 m*M* $\beta$-mercaptoethanol, 5%(*v*/*v*) glycerol and 10 m*M* imidazole] and eluted using a 0–1 *M* imidazole gradient. SurE protein eluted near 500 m*M* imidazole, as confirmed by SDS–PAGE. Pooled fractions were exchanged into a final buffer (50 m*M* sodium succinate pH 5.0, 100 m*M* NaCl) using PD10 columns (Amersham-Pharmacia) and concentrated to 9 mg ml$^{-1}$. A uniform population of SurE was confirmed by gel filtration with Superdex-75 (Amersham-Pharmacia) using the same buffer. The final yield of pure protein was estimated at 40 mg per litre of culture broth. Protein concentrations were determined using the Bradford assay (Bradford, 1976).

### 2.2. Crystallization of SurE

Automated crystallization screening was carried out at 295 K using sitting-drop vapour diffusion in 96-well plates (Greiner) using a Cartesian Minibee nanolitre pipetting robot (Genomic Solutions) and screens from Nextal Biotechnologies (Classics, PEGs, pH Clear, MbClass and MbClass II). Each crystallization drop contained 150 nl protein solution and 150 nl reservoir solution and was equilibrated against 150 µl reservoir solution.

Two distinct crystal forms were reproducibly obtained. Crystallization drops were set up at 294 K using the hanging-drop vapour-diffusion method. In one condition (condition 1) crystals appeared within 48 h in 0.2 *M* magnesium acetate and 20%(*w*/*v*) polyethylene

glycol (PEG) 3350. The quality of these crystals was further optimized by adding lithium chloride to the drop to a final concentration of 0.01 *M*. The second condition, 0.02 *M* sodium citrate, 0.1 *M* sodium dihydrogen phosphate pH 6.2 and 15%(*w*/*v*) polythylene glycol 2000 (condition 2), produced crystals in 24–48 h. In the optimization process for both conditions, crystals were grown by preparing drops with 1 µl protein solution and 1 µl crystallization solution and equilibrating against 0.5 ml of the same solution in the reservoir (Fig. 1). The pH quoted in Table 1 for crystal growth is the overall pH of the reservoir solution prior to mixing with the protein solution. The crystal from condition 1 (crystal form 1) belonged to the monoclinic space group $P2_12_12_1$, while the crystal form derived from condition 2 (crystal form 2) grew in the monoclinic space group *C*2. Table 1 contains full details of all crystallization conditions and unit-cell parameters.

### 2.3. Data collection

Crystals from condition 1 were flash-cooled in liquid nitrogen using the optimized crystallization solution modified by a 2% increment in the precipitant concentration and the addition of 20%(*v*/*v*) PEG 400. The cryoprotectant for the crystal from condition 2 was optimized by sequentially transferring the crystal to the optimized crystallization solution with an increased polythylene glycol 2000 content [17%(*w*/*v*)] and progressively increasing amounts of glycerol up to a final value of 20%(*v*/*v*). Following this procedure, the crystals were flash-cooled in liquid nitrogen. The data sets were collected at 100 K using an Oxford Cryostream setup on the ID14-EH2 and ID23-EH2 synchrotron beamlines (ESRF, Grenoble, France) using an ADSC Quantum 4R CCD detector and a MAR Mosaic 225 CCD detector, respectively. Data processing and scaling were performed using *MOSFLM* (Leslie, 2006) and *SCALA* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Full data-processing statistics and parameters are shown in Table 1.

**Table 1**
Crystallization conditions and diffraction data of *Cj*SurE.

Values in parentheses are for the outer shell.

| | Crystal form 1 | Crystal form 2 |
|---|---|---|
| Buffer | — | 0.1 m*M* sodium dihydrogen phosphate pH 6.2 |
| Salt | 0.2 *M* magnesium acetate | 0.02 *M* sodium citrate |
| Precipitant | 20%(*w*/*v*) PEG 3350 | 15%(*w*/*v*) PEG 2000 |
| Overall pH | 7.4 | 6.5 |
| Beamline | ID14-2 | ID23-2 |
| Space group | $P2_12_12_1$ | *C*2 |
| Solvent content (%) | 55.80 | 49.28 |
| $V_M$ (Å$^3$ Da$^{-1}$) | 2.78 | 2.42 |
| Molecules per ASU | 4 | 2 |
| Unit-cell parameters | | |
| *a* (Å) | 80.5 | 121.4 |
| *b* (Å) | 119.0 | 47.1 |
| *c* (Å) | 135.3 | 97.8 |
| $\alpha$ (°) | 90.0 | 90.0 |
| $\beta$ (°) | 90.0 | 95.7 |
| $\gamma$ (°) | 90.0 | 90.0 |
| Resolution limits | 35.53–2.50 (2.65–2.50) | 49.30–2.85 (2.95–2.85) |
| Wavelength (Å) | 0.933 | 0.873 |
| No. of measurements | 318696 | 41796 |
| No. of unique measurements | 44383 | 13085 |
| Completeness (%) | 99.8 (96.7) | 98.2 (94.8) |
| Mean $I/\sigma(I)$ | 20.1 (2.9) | 13.0 (2.3) |
| $R_{merge}$† (%) | 6.9 (27.1) | 6.9 (29.7) |
| $R_{p.i.m.}$‡ (%) | 2.7 (14.2) | 4.6 (15.5) |
| Mean temperature *B* factor (Å$^2$) | 45.8 | 55.3 |
| Correlation coefficient after MR | 0.48 | 0.41 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean intensity of reflection *hkl*. ‡ $R_{p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where *N* is the redundancy of a given reflection.

## 3. Results and discussion

Recombinant *C. jejuni* stationary-phase survival E protein was successfully expressed in *E. coli*, purified and crystallized in two distinct crystal forms. Dynamic light scattering and gel-filtration

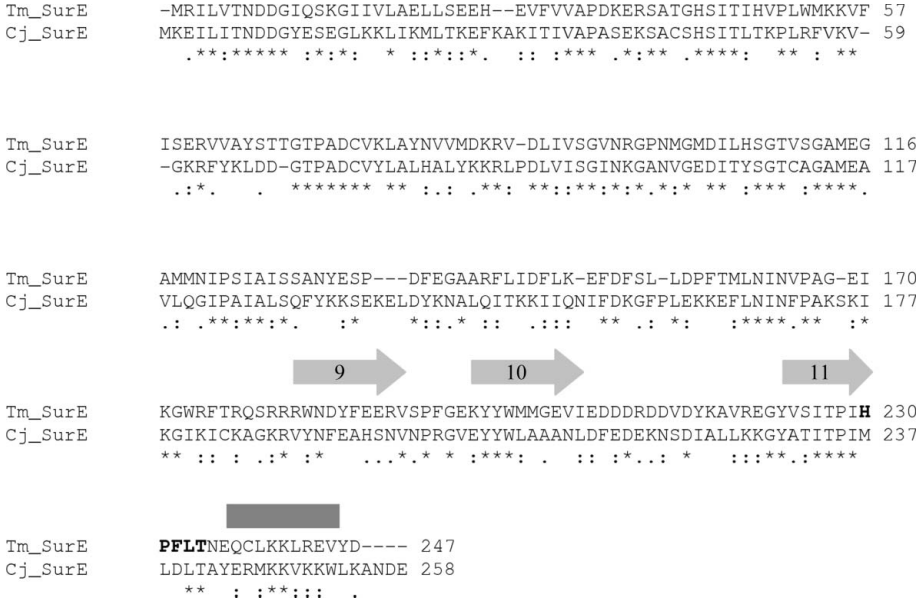



**Figure 2**
Sequence alignment of *Tm*SurE and *Cj*SurE proteins. Amino-acid identities are marked '*', with similar residues marked '.' and very similar residues marked ':'. Secondary-structure elements (in *Tm*SurE) are only indicated in the C-terminal region. $\beta$-Sheets 9 and 10 (light grey arrows) are involved in the tetramerization of *Tm*SurE. The hinge loop (residues in bold) is flanked by a highly conserved $\beta$-sheet ($\beta$11) and a C-terminal domain-swapped $\alpha$-helix (dark grey rectangle). This alignment was produced with *ClustalW* (Thompson *et al.*, 1994).

chromatography suggest that *Cj*SurE exists as a dimer in solution. The crystal forms obtained belonged to space groups *C*2 and $P2_12_12_1$ and revealed the likely presence of a dimer and a tetramer, respectively, in the asymmetric unit. Initial data collection from the *C*2 form was affected by the high mosaicity of the crystals. This was partially overcome by a slow and gradual increase in glycerol concentration during pre-cooling soaks in order to allow the crystal to accommodate and resist the osmotic pressure of the cryoprotectant solution.

Previous structural studies of *Tm*SurE have indicated domain swapping involving the C-terminal $\alpha$-helix (Lee *et al.*, 2001; Zhang *et al.*, 2001). However, comparative sequence analysis of *Cj*SurE and *Tm*SurE also indicates high sequence similarity in the N-terminal domains (Fig. 2) of these proteins. Although the divergence between these enzymes is greater in the C-terminal region, the structural features determined in the crystal structure of *Tm*SurE that are believed to mediate tetramerization of this protein, predominantly a C-terminal $\beta$-hairpin ($\beta$9 and $\beta$10), may also be present in *Cj*SurE. The final 25–30 amino acids in *Cj*SurE and *Tm*SurE contain a highly conserved stretch of amino acids corresponding to a final $\beta$-sheet ($\beta$11), which is followed by a short loop hinge (residues in bold in Fig. 2) that allows the C-terminal $\alpha$-helix domain swap in *Tm*SurE. The biological significance of this feature is unknown.

This research benefitted from funding from the European Commission under the SPINE project (contract No. QLG2-CT-2002-00988 under the RTD programme 'Quality of Life and Management of Living Resources'). The authors would like to thank Dr Edward Mitchell for his assistance in collecting data on ID14-EH2 (ESRF). Special thanks is given to Dr Daniele de Sanctis for help during data collection on ID23-EH2 (ESRF). AMDG acknowledges Fundação para a Ciência e Tecnologia for a postdoctoral research grant (SFRH/BPD/14553/2003).

## References

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Galperin, M. Y., Moroz, O. V., Wilson, K. S. & Murzin, A. G. (2006). *Mol. Microbiol.* **59**, 5–19.
Iwasaki, W. & Miki, K. (2007). *J. Mol. Biol.* **371**, 123–136.
Kamiya, H. (2003). *Nucleic Acids Res.* **31**, 517–531.
Lee, J. Y., Kwak, J. E., Moon, J., Eom, S. H., Liong, E. C., Pedelacq, J.-D., Berendzen, J. & Suh, S. W. (2001). *Nature Struct. Biol.* **8**, 789–794.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Li, C., Wu, P. Y. & Hsieh, M. (1997). *Microbiology*, **143**, 3513–3520.
Loewen, P. C. & Hengge-Aronis, R. (1994). *Annu. Rev. Microbiol.* **48**, 53–80.
Mura, C., Katz, J. E., Clarke, S. G. & Eisenberg, D. (2003). *J. Mol. Biol.* **326**, 1559–1575.
Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N. N., Kitagawa, M., Mori, H., Savchenko, A. & Yakunin, A. (2004). *J. Biol. Chem.* **279**, 54687–54694.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Visick, J. E., Cai, H. & Clarke, S. (1998). *J. Bacteriol.* **180**, 2623–2629.
Zhang, R.-G., Skarina, T., Katz, J. E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C. H., Clarke, S., Edwards, A., Joachimiak, A. & Savchenko, A. (2001). *Structure*, **9**, 1095–1106.