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

Crystallization and preliminary X-ray analysis of the mRNA-binding domain of elongation factor SelB from *Escherichia coli* in complex with RNA

Nicolas Soler, Dominique Fourmy and Satoko Yoshizawa*

Laboratoire de Chimie et Biologie Structurales, ICSN–CNRS, 1 Avenue de la Terrasse, 91190 Gif-sur-Yvette, France

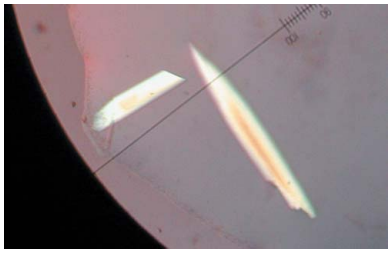
Correspondence e-mail: yoshizawa@icsn.cnrs-gif.fr

Received 13 February 2007
Accepted 6 April 2007

In bacteria, selenocysteine (the 21st amino acid) is incorporated into proteins *via* machinery that includes SelB, a specific translational elongation factor. SelB binds to an mRNA hairpin called the selenocysteine-insertion sequence (SECIS) and delivers selenocysteyl-tRNA^{Sec} to the ribosomal A site. The minimum C-terminal fragment (residues 478–614) of *Escherichia coli* SelB (SelB-WH3/4) required for SECIS binding has been overexpressed and purified. This protein was crystallized in complex with 23 nucleotides of the SECIS hairpin at 294 K using the hanging-drop vapour-diffusion method. A data set was collected to 2.3 Å resolution from a single crystal at 100 K using ESRF beamline BM-30. The crystal belongs to space group *C*2, with unit-cell parameters $a = 103.50$, $b = 56.51$, $c = 48.41$ Å. The asymmetric unit contains one WH3/4-domain–RNA complex. The Matthews coefficient was calculated to be 3.37 Å^3 Da^{-1} and the solvent content was estimated to be 67.4%.

1. Introduction

Selenocysteine (the 21st amino acid) is present in all three kingdoms of life. In bacteria, its incorporation into proteins requires highly specific machinery including elongation factor SelB (Forchhammer *et al.*, 1989). SelB is able to interact with the ribosome, selenocysteyl-tRNA^{Sec} and guanine nucleotides *via* its N-terminal part, which is homologous to the three domains of elongation factor Tu (EF-Tu). An extra C-terminal domain (SelB-C) binds an mRNA hairpin called SECIS (selenocysteine-insertion sequence), which serves as a signal to deliver Sec-tRNA^{Sec} to a UGA codon at the ribosomal A site. Hence, a UGA codon with a downstream SECIS is recoded as a Sec codon instead of being used as a termination signal. The crystal structure of SelB-C from *Moorella thermoacetica* (Selmer & Su, 2002) and the solution structure of the *Escherichia coli* SECIS hairpin (Fourmy *et al.*, 2002) have been reported. SelB-C is composed of four consecutive winged-helix motifs (WH motifs), a module that is found in many DNA-binding proteins (Gajiwala & Burley, 2000). The crystal structure of the *M. thermoacetica* SelB WH3/4 C-terminal domain in complex with SECIS (Yoshizawa *et al.*, 2005) has shed light on the molecular basis of this interaction. The WH4 domain recognizes the extruded nucleotides G23 and U24 at the tip of the SECIS stem loop as well as five consecutive helical phosphates on the 5′ side of the hairpin loop. However, the affinity between *M. thermoacetica* SelB and SECIS is in the micromolar range (Yoshizawa *et al.*, 2005), which is three orders of magnitude lower than in the case of *E. coli* SelB (nanomolar range; Thanbichler *et al.*, 2000). This observation raises the possibility that the high affinity of *E. coli* SelB for mRNA originates from the use of the modular organization of the WH motifs. This is supported by the identification of a bulged uracil residue (U17) located outside the main protein-binding site as being crucial for high-affinity binding of mRNA in *E. coli* SECIS (Hüttenhofer *et al.*, 1996; Liu *et al.*, 1998). In the previous crystal structure of a WH3/4–SECIS RNA complex from *M. thermoacetica*, the predicted position of the bulged uracil lies in the vicinity of the WH3 domain (Yoshizawa *et al.*, 2005). Thus, the WH3 module in *E. coli* SelB may directly establish specific contacts with SECIS RNA.



In this paper, we report the production and crystallization of the minimum fragment of the RNA-binding domain (SelB-WH3/4; residues 478–614) of SelB from *E. coli* in complex with the SECIS mRNA hairpin. This will allow the investigation of the RNA-binding activity of the two consecutive WH motifs (WH3-WH4) in *E. coli* SelB, for which abundant genetic, biochemical and kinetic data are available.

2. Results and discussion

2.1. Cloning, expression and purification of SelB-WH3/4 and complex formation with SECIS RNA

The sequence of the *E. coli* SelB C-terminal fragment (residues 468–614) was amplified by PCR from plasmid DNA containing a slightly larger sequence of the same domain (Fourmy *et al.*, 2002) and cloned into the plasmid vector pGEX-2T (Amersham Biosciences) for overexpression as a GST-fusion protein. The protein was overexpressed and purified from *E. coli* BL21. The fusion protein was purified essentially as described by Smith & Johnson (1988), with slight modifications. The harvested cells were lysed by sonication and centrifuged. The supernatant was loaded onto a glutathione agarose column (Sigma). The GST-fusion protein was eluted with 5 m*M* reduced glutathione (Sigma) in 50 m*M* Tris–HCl pH 8.0. The GST was removed from the SelB C-terminal fragment by cleavage with thrombin (Sigma) and purification took place using glutathione agarose column chromatography followed by CM-Sepharose column chromatography (Amersham Biosciences). To remove the potentially unstructured N- or C-terminal tail that may be unfavourable for crystallization, the purified protein was subjected to chymotrypsin cleavage for 1 h at 310 K (Fig. 1, lane 4). Mass spectrometry allowed the identification of the N-terminal cleavage site as being after residue Trp477, which left a new SelB C-terminal fragment (SelB-WH3/4; residues 478–614). The fragment cleaved by chymotrypsin corresponds to the C-terminal sequence of winged-helix motif 2, which is unstructured. SelB-WH3/4 was then cloned into pGEX-2T using primers 5′-CGCGGATCCCTGCATCTGCCAGATCACAAAGCG-3′ and 5′-CGTAATAATAAAGGCCTTTTTATTCTTAAGCCC-3′ and overexpressed in *E. coli* BL21 using the same protocol as for the larger fragment. The 23-nucleotide SECIS RNA hairpin (5′-GGCGGUUGCAGGUCUGCACCGCC-3′) derived from the *E. coli fdhF* gene was transcribed *in vitro* and purified by denaturing polyacrylamide gel electrophoresis (PAGE) as described by Fourmy *et al.* (2002). Two 24-nucleotide SECIS RNA constructs containing a single G or C nucleotide 3′-overhang were produced and screened for their crystallization properties (as an equimolar mixture of the two constructs). These two nucleotides are complementary and often make a base pair in crystals of protein–DNA or protein–RNA complexes. Two pairs of 26- and 22-nucleotide RNA constructs containing a single G or C nucleotide 3′-overhang were also tested in order to increase and decrease the stem length by one base pair. All these constructs failed to provide crystals that were suitable for structure determination. This may be explained from the obtained crystal structure (Soler, Fourmy and Yoshizawa, submitted): the SECIS helices are coaxially stacked, with a lateral shift at the helical termini. Prior to crystallization, the 23-nucleotide SECIS RNA hairpin was heated at 368 K for 1 min and cooled to 277 K on ice. Equimolar amounts of SelB-WH3/4 protein and SECIS RNA were mixed and the SelB-WH3/4–SECIS RNA complex was concentrated to a final concentration of 6.75 mg ml^{-1} using a Centricon YM-3 filter (Millipore). Finally, the complex was dialyzed against 10 m*M* Tris–HCl pH 7.0, 100 m*M* NaCl, 0.1 m*M* EDTA. The complex solution was stored at 277 K until use.

2.2. Crystallization of the SelB-WH3/4–SECIS RNA complex

Crystals of the complex were obtained by hanging-drop vapour diffusion at 294 K using 24-well Linbro plates (Hampton Research). Typically, a hanging drop was prepared by mixing equal volumes (1 µl each) of protein–RNA complex solution (at 6.75 mg ml^{-1}) and reservoir solution. Each hanging drop was placed over 0.7 ml reservoir solution. Preliminary crystallization trials were conducted at 294 K using Crystal Screen I and Natrix solution kits (Hampton Research). Crystals appeared within a few days for the SelB C-terminal fragment using Crystal Screen I condition No. 1 [CSI-1; 0.02 *M* calcium chloride dihydrate, 0.1 *M* sodium acetate trihydrate pH 4.6, 30%(*v*/*v*) 2-methyl-2,4-pentanediol] but were not stable over time, disappearing after 1 or 2 d. As mentioned above, chymotrypsin cleavage combined with mass spectrometry was used to identify and clone a slightly shorter and more structured version of the SelB C-terminal fragment (SelB-WH3/4). Crystals of SelB-WH3/4 were obtained with CSI-1. The resolution limit of the crystals was typically between 4.0 and 2.6 Å. Additives were screened in order to optimize the crystallization conditions. All conditions from Additive Screens I, II and III (Hampton Research) were tested by mixing equal volumes of complex, reservoir and Additive Screen stock solutions (1 µl each). Crystals obtained with CSI-1 and Additive Screen III condition No. 6 (ASIII-6; 0.5 *M* NaF) were suitable for structure determination. Needle-shaped crystals appeared within 2 d and reached dimensions of 40 × 40 × 100–500 µm) after two weeks (Fig. 2).

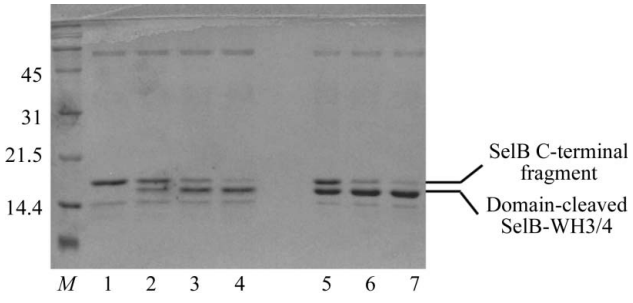


Figure 1
Chymotrypsin cleavage of the SelB C-terminal fragment. Lane *M*: molecular-weight standards (kDa). Lane 1: purified SelB C-terminal fragment. Lanes 2, 3 and 4: 10, 30 and 60 min after the addition of chymotrypsin [1/50(*w*/*w*) with respect to SelB C-terminal domain], respectively. Lanes 5, 6 and 7: RNA complex 10, 30 and 60 min after the addition of chymotrypsin [1/50(*w*/*w*)], respectively.

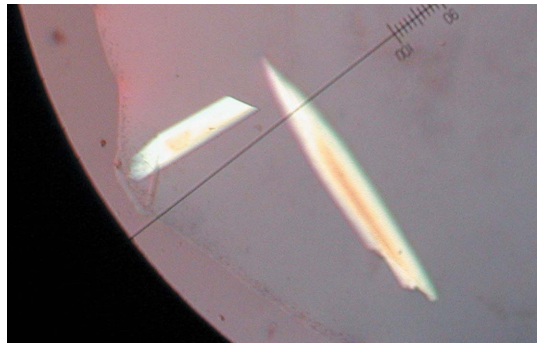


Figure 2
Crystals of the SelB-WH3/4–SECIS RNA complex in a hanging drop. Crystal dimensions are 40 × 40 × 200 µm.

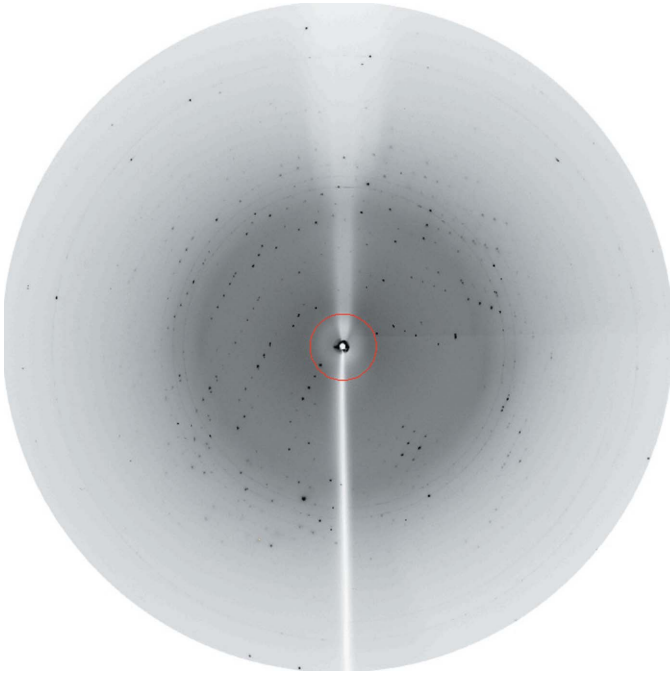


Figure 3
X-ray diffraction image of a SelB-WH3/4–SECIS RNA complex crystal. The detector edge corresponds to 1.95 Å resolution. The exposure time was 30 s, the detector distance was 150 mm and the oscillation range per frame was 1°.

2.3. Preliminary X-ray diffraction analysis

For diffraction data collection, crystals of the SelB-WH3/4–SECIS RNA complex obtained using CSI-1 and ASIII-6 were directly soaked in a cryoprotectant solution consisting of CSI-1 containing 15% ethylene glycol. A data set was collected to 2.3 Å (Fig. 3) from a single crystal at 100 K on beamline BM30 at ESRF (Grenoble, France) (λ = 0.980 Å). The diffraction data were collected using 1° oscillations with a crystal-to-detector distance of 150 mm. The diffraction data were autoindexed, integrated with *MOSFLM* (Leslie, 1992) and scaled with *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The data-collection statistics are summarized in Table 1. Crystals belong to space group *C*2, with unit-cell parameters a = 103.50, b = 56.51, c = 48.41 Å, α = 90, β = 95.60, γ = 90° ($R_{\rm merge}$ = 8% on data to 2.3 Å resolution, data completeness = 96.3%), and contained one complex per asymmetric unit. The structure was solved by molecular replacement at 2.3 Å resolution using the program *Phaser* with the *M. thermoacetica* SelB-WH3/4–SECIS complex (Yoshizawa *et al.*, 2005) as a search model and was refined using *REFMAC* (Collaborative Computational Project, Number 4, 1994). The detailed structural description will be published elsewhere (Soler, Fourmy and Yoshizawa, submitted). The PDB code is 2pjp.

Table 1
Data-collection statistics.

Wavelength (Å)	0.980
Resolution range (Å)	18–2.30 (2.42–2.30)
Measured reflections	43410 (6296)
Unique reflections	11966 (1737)
Completeness (%)	96.3 (96.2)
Mean $I/\sigma(I)$	17.4(2.6)
$R_{\rm merge}$ (%)	8.0 (51.8)

We thank the FIP group and in particular Lilian Jacquamet for assistance in data collection at ESRF Grenoble. This work was supported by an ANR Young Researcher grant to DF and SY.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Forchhammer, K., Leinfelder, W. & Böck, A. (1989). *Nature (London)*, **342**, 453–456.
Fourmy, D., Guittet, E. & Yoshizawa, S. (2002). *J. Mol. Biol.* **324**, 137–150.
Gajiwala, K. S. & Burley, S. K. (2000). *Curr. Opin. Struct. Biol.* **10**, 110–116.
Hüttenhofer, A., Westhof, E. & Böck, A. (1996). *RNA*, **2**, 354–366.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Liu, Z., Reches, M., Groisman, I. & Engelberg-Kulka, H. (1998). *Nucleic Acids Res.* **26**, 896–902.
Selmer, M. & Su, X.-D. (2002). *EMBO J.* **21**, 4145–4153.
Smith, D. B. & Johnson, K. S. (1988). *Gene*, **67**, 31–40.
Thanbichler, M., Böck, A. & Goody, R. S. (2000). *J. Biol. Chem.* **275**, 20458–20466.
Yoshizawa, S., Rasubala, L., Ose, T., Kohda, D., Fourmy, D. & Maenaka, K. (2005). *Nature Struct. Mol. Biol.* **12**, 198–203.