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

Krishan Gopal Thakur and B. Gopal*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: bgopal@mbu.iisc.ernet.in

Received 2 June 2005
Accepted 15 July 2005
Online 30 July 2005

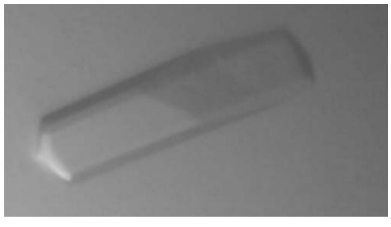
Crystallization and preliminary X-ray diffraction studies of two domains of a bilobed extra-cytoplasmic function sigma factor SigC from *Mycobacterium tuberculosis*

Sigma factors are transcription-regulatory proteins that bind to RNA polymerase and facilitate promoter recognition. The so-called extracytoplasmic function sigma factors help a bacterium to respond to environmental conditions. *Mycobacterium tuberculosis* SigC (σ^{C}) is an extracytoplasmic sigma factor that is essential for lethality in a mouse model of infection and is conserved in all pathogenic mycobacterial species. This protein consists of two domains that are connected by an ~25-amino-acid linker. The N-terminal domain contains the σ_2 DNA-binding motif, whereas the σ_4 motif is located in the C-terminal domain. Native σ^{C} did not yield diffraction-quality crystals. However, two of its domains have been cloned, expressed and crystallized: σ_2^{C} (12.3 kDa) and σ_4^{C} (7.5 kDa). The σ_2^{C} crystals belong to the hexagonal space group $P6_1$, with unit-cell parameters $a = b = 85.28$, $c = 79.63$ Å, and native X-ray diffraction data were collected from this domain to 2.7 Å on an in-house X-ray home source. The σ_4^{C} crystals belong to the cubic space group $F23$, with unit-cell parameters $a = b = c = 161.21$ Å. X-ray diffraction data were collected from this domain to 3.1 Å, also on an in-house X-ray source.

1. Introduction

DNA-dependent RNA polymerase (RNAP), the main component of bacterial gene expression, consists of five subunits. The two α-subunits together with the β- and β'-subunits form a stable catalytic complex that is involved in RNA elongation. The fifth subunit of RNA polymerase is a sigma factor that provides promoter-recognition features to the RNA polymerase and is thus essential for the initiation of transcription. *Mycobacterium tuberculosis* has a total of 13 sigma (σ) factors: three primary sigma factors, ten extra-cytoplasmic function (ECF) sigma factors and one alternative sigma factor (pvdS). The difference in the promoter specificity of these sigma factors leads to differential gene expression. Regulatory mechanisms dictate which sigma factors are activated and bind to the apo-RNA polymerase and initiate transcription. This sigma-regulatory protein mechanism is coupled to a signal transduction system that senses the environment. The resulting interplay between signal transduction and the transcriptional regulatory mechanisms allows the bacillus to respond to changes in the environment, inducing or downregulating the synthesis of proteins. This ability is essential for the bacillus to survive the stresses inside the host until a failure in the host defences leads to a reactivation of the disease.

The ECF σ factors belong to the σ^{70} family of transcriptional regulators. Many of the ECF sigma factors are negatively regulated by the antisigma (anti-σ) factor(s) and control the expression of a relatively small number of regulons (Lonetto *et al.*, 1994). Although there are four conserved promoter-recognition motifs (regions 1–4) in the σ^{70} family of sigma factors, only regions 2 and 4 are maintained in the ECF σ factors. Multiple roles have been assigned to region 2 of the sigma factor; this region has been shown to be involved in core RNAP binding, DNA melting and −10 DNA (Pribnow box) recognition (Gross *et al.*, 1998; Lonetto *et al.*, 1994). Region 4 of the σ factor has been shown to interact with the −35 region of promoter DNA (Gross *et al.*, 1998; Campbell *et al.*, 2003). *M. tuberculosis* σ^{C} shares a substantial similarity with *E. coli* σ^{E} (28.1% identity between the two proteins and ~33% identity between residues 20–180). The structure



of *Escherichia coli* σ^{E} bound to its anti-σ factor RseA has been determined (Campbell *et al.*, 2003). The crystal structure of σ^{E} revealed two independently folded globular α-helical domains connected by an ~25-residue linker region. On the basis of sequence alignment with *E. coli* σ^{E}, the two domains of σ^{C} have been assigned as σ_2^{C} (N-terminal domain containing the conserved region 2 that spans residues 1–112) and σ_4^{C} (C-terminal domain containing the region 4 and comprised of residues 117–185). Although evidence about the role of σ^{C} and the mechanism of its activation is very limited, the conservation of this protein across all pathogenic mycobacterial species suggests a potential role for this protein in pathogenesis. The protein σ^{C} has been shown to be required for lethality in a mouse model of mycobacterial infection. For example, the mycobacterium σ^{C}-knockout phenotype in time-to-death experiments in the mouse model was significantly attenuated, causing no death in infected mice (Sun *et al.*, 2004). It has also been found that σ^{C} modulates the expression of several key virulence-associated genes including *hspX*, *senX3* and *mtrA* (Sun *et al.*, 2004).

In view of the bilobed nature of bacterial sigma factors, it is not surprising that all intact structures of σ factors known to date are either in complex with their respective antisigma factor or bound to RNAP. Structures of the domains containing the conserved regions of σ factors have also been determined; they do not suggest major changes in the domain structure upon complex formation (Malhotra *et al.*, 1996; Campbell *et al.*, 2002). Crystallization trials on intact full-length σ^{C} led to crystalline precipitates under some conditions, but these could not be improved any further. To circumvent the problem (which is likely to arise from conformational heterogeneity caused by the flexible linker connecting the two domains), we subcloned the two globular domains of σ^{C}: σ_2^{C} (MW 12.3 kDa) and σ_4^{C} (MW 7.3 kDa), expressed, purified and crystallized them separately.

2. Materials and methods

2.1. Expression and purification of the recombinant proteins

Three different expression constructs were used in this study: the intact σ^{C} (Rv2069) and the two domains σ_2^{C} and σ_4^{C}. The gene encoding σ^{C} (residues 1–185) and σ_4^{C} (residues 117–185) were cloned between the *Nde*I and *Bam*HI restriction sites of bacterial expression vector pET-15b (Novagen), whereas the σ_2^{C} region of σ^{C} (residues 1–112) was cloned in a modified version of the pET-15b vector utilizing the *Nhe*I and *Xho*I restriction sites. After transforming the plasmid into BL21 (DE3) cells (Novagen Inc.), the cells were grown in Luria broth with ampicillin (100 µg ml^{-1}) to an $OD_{600\,nm}$ of 0.5–0.6. The cells were induced with 0.2 m*M* IPTG (final concentration). Subsequently, the growth temperature was lowered to 290 K and cells were grown for a further 12 h before they were spun down and stored at 193 K. The cells were resuspended in buffer *A* (50 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5) and lysed using sonication (three bursts of 1 min each at medium setting) on ice. The cell debris was separated from the crude cell lysate by centrifugation at 26 892*g* for 30 min at 277 K in a Sorvall centrifuge. The cell-free lysate was equilibrated with Ni-CAM resin (Sigma-Aldrich Inc.; approximately 5 ml resin suspension was used for the cell-free lysate from 5 g cell paste). The protein was eluted from the Ni-CAM column using a gradient of imidazole concentration (10–200 m*M*) in buffer *A*. The purified protein (Fig. 1) was further subjected to size-exclusion chromatography on a Sephacryl Hi-Prep 16/60 S-200 HR column (Amersham Biosciences, Inc.) equilibrated with buffer *A* at a flow rate of 0.5 ml min^{-1}. The purity of the sample was analyzed using SDS–PAGE followed by Coomassie blue staining. The molecular weights of the recombinant proteins were also examined by mass spectrometry on a MALDI–TOF (Bruker Daltonics) mass spectrometer.

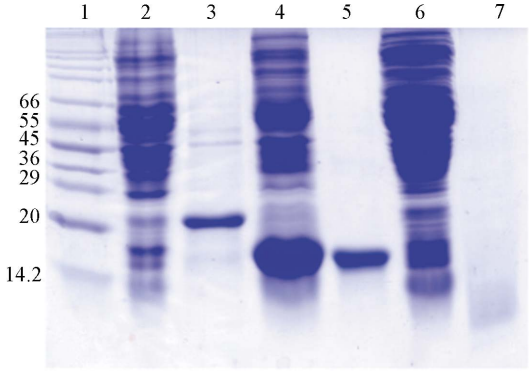


Figure 1
A 12% SDS–PAGE showing the overexpression and purification of σ^{C} constructs. Lane 1, markers (molecular weights in kDa); lane 2, σ^{C} lysate; lane 3, σ^{C} purified (22.1 kDa); lane 4, σ_2^{C} lysate; lane 5, purified σ_2^{C} (14.5 kDa); lane 6, σ_4^{C} lysate; lane 7, purified σ_4^{C} (9.5 kDa).

2.2. Crystallization and data collection

Initial screening for crystallization conditions for the proteins was performed using crystallization kits from Hampton Research (Crystal Screens 1 and 2, Index Screen and the PEG/Ion Screen). The purified proteins (σ_2^{C} and σ_4^{C}) were concentrated to ~3 mg ml^{-1} in 50 m*M* Tris–HCl buffer containing 250 m*M* NaCl, 50 m*M* L-arginine, 50 m*M* L-glutamic acid and 100 m*M* imidazole pH 7.5. These conditions were examined using the hanging-drop method at 293 K in which the drop contained 2 µl protein solution and 2 µl crystallization solution. Crystalline precipitates were observed in several conditions for intact σ^{C}, but they could not be improved. However, the two separate domains were readily crystallizable. σ_2^{C} crystallized in a condition containing 2 *M* ammonium sulfate and 5% 2-propanol. Although σ_4^{C} crystallized under several conditions, the condition that gave the best crystals consisted of 1.8 *M* ammonium sulfate, 0.1 *M* MES pH 6.5, 1 m*M* DTT and 5% dioxane. Variations around these crystallization conditions were examined by changing the concentrations of ammonium sulfate and the organic solvent. These variations resulted in only minor improvement in crystal size for the two domains, which were observed within 2–3 d of setting up the hanging-drop experiment. The cryoprotection protocols for σ_2^{C} and σ_4^{C} crystals were rather similar as both required 20% glycerol as the cryoprotectant. The gradient of increase in the cryoprotectant concentration to achieve the final concentration of 20%(*v*/*v*) was optimized to provide the lowest mosaicity upon cooling. The flash-cooling protocol involved the transfer of the crystals into buffers containing 10, 15 and 20% glycerol concentration. The crystals were soaked for 1 min in

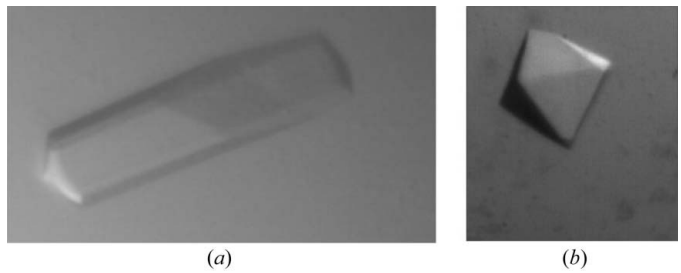


Figure 2
(*a*) Hexagonal crystals (space group $P6_1$) of σ_2^{C} obtained in the presence of ammonium sulfate and 2-propanol. (*b*) Crystal of σ_4^{C} belonging to cubic space group *F*23 obtained in the presence of ammonium sulfate and dioxane.

Table 1
Summary of data-collection, processing and refinement statistics for σ_4^C and σ_4^C.

Values in parentheses are for the outer shell.

Parameters	σ_2^C (N-terminal domain of sigC)	σ_4^C (C-terminal domain of sigC)
Wavelength (Å)	1.5418	1.5418
Resolution limits (Å)	35.01–2.70 (2.89–2.70)	40.29–3.10 (3.27–3.1)
Total No. of observations	45969 (6060)	38587 (4717)
Total No. of unique reflections	17709 (2612)	6420 (939)
Completeness (%)	99.8 (99.2)	100.0 (99.9)
Multiplicity	2.6 (2.3)	6.0 (5.0)
R_{merge}†	0.087 (0.387)	0.110 (0.489)
$\langle I\rangle/\sigma(I)$	10.9 (2.3)	15.8 (3.3)

† $R_{merge} = \sum_j |\langle I\rangle - I_j|/\sum I$, where I_j is the intensity of the jth reflection and $\langle I\rangle$ is the average intensity.

each cryotransfer solution. The crystals were flash-cooled in gaseous N_2 (100 K). The diffraction data were collected at 100 K on a MAR imaging-plate system mounted on a Rigaku RU-200 rotating-anode X-ray generator. The data were processed using *MOSFLM* (Leslie, 1992) and scaled using the program *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

The crystals of both the σ_2^C and σ_4^C domains appeared within 2–3 d of setting up the hanging-drop experiment and grew to maximum dimensions within a week (~0.4 × 0.08 × 0.08 mm for σ_2^C and ~0.1 × 0.1 × 0.1 mm for σ_2^C; Fig. 2). σ_2^C crystals diffracted to better than 2.7 Å resolution and a complete data set could be collected from a single crystal. The σ_4^C crystals diffracted to better than 3.1 Å. The data-collection statistics for both the domains are reported in Table 1. Based on molecular weight, space group and self-rotation function, the σ_2^C crystal was seen to contain two molecules in the asymmetric unit, with a V_M value of 2.8 Å^3 Da^{-1}, corresponding to a solvent content of 55.7% (there is no evidence of the functional oligomerization of ECF sigma factors). The N-terminal domain of *E. coli* σ^E (PDB code 1or7; Campbell *et al.*, 2003) served as the initial model (sequence identity 23%) for solving the structure by molecular replacement using *PHASER* (Read, 2001; Storoni *et al.*, 2004). The solution had a log-likelihood gain of 71.52. The σ_4^C crystal contained three molecules in the asymmetric unit, giving a V_M value of 2.9 Å^3 Da^{-1} with 57.3% solvent content for σ_4^C. *MOLREP* (Vagin & Teplyakov, 1997) was used with the C-terminal domain of *E. coli* σ^E (sequence identity ~45%) as a model (PDB code 1or7; Campbell *et al.*, 2003) to locate three molecules in the asymmetric unit. The molecular-replacement solution for σ_4^C had a correlation coefficient of 0.31 and an R factor of 0.56. Model building and refinement of these structures are in progress.

The work reported in this manuscript was funded by a grant from the Council for Scientific and Industrial Research, Government of India to BG.

References

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002). *Mol. Cell*, **9**, 527–539.
Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. & Darst, S. A. (2003). *Mol. Cell*, **11**, 1067–1078.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. & Young, B. (1998). *Cold Spring Harbor Symp. Quant. Biol.* **63**, 141–155.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lonetto, M. A., Brown, K. L., Rudd, K. E. & Buttner, M. J. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 7573–7577.
Malhotra, A., Severinova, E. & Darst, S. A. (1996). *Cell*, **87**, 127–136.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Sun, R., Converse, P. J., Ko, C., Tyagi, S., Morrison, N. E. & Bishai, W. R. (2004). *Mol. Microbiol.* **52**, 25–38.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.