Crystallization and preliminary X-ray crystallographic analysis of carboxyl-terminal region 4 of SigR from *Streptomyces coelicolor* A3(2)

Keon Young Kim, Sunmin Kim,
Jeong Kuk Park, Hyojin Song and
SangYoun Park*

School of Systems Biomedical Science, Soongsil
University, Seoul 156-743, Republic of Korea

Correspondence e-mail: psy@ssu.ac.kr

Received 24 January 2014

Accepted 14 April 2014

Full-length SigR from *Streptomyces coelicolor* A3(2) was overexpressed in *Escherichia coli*, purified and submitted to crystallization trials using either polyethylene glycol 3350 or 4000 as a precipitant. X-ray diffraction data were collected to 2.60 Å resolution under cryoconditions using synchrotron X-rays. The crystal packs in space group $P4_32_12$, with unit-cell parameters $a = b = 42.14$, $c = 102.02$ Å. According to the Matthews coefficient, the crystal asymmetric unit cannot contain the full-length protein. Molecular replacement with the known structures of region 2 and region 4 as independent search models indicates that the crystal contains only the −35 element-binding carboxyl-terminal region 4 of full-length SigR. Mass-spectrometric analysis of the harvested crystal confirms this, suggesting a crystal volume per protein weight (V_M) of $2.24 \text{ Å}^3 \text{ Da}^{-1}$ and 45.1% solvent content.

1. Introduction

Bacterial gene expression at the level of transcription is initiated by alternative forms of RNA polymerases (RNAPs) that recognize their specific promoters. Moreover, the affinity between a certain promoter and an RNAP is further enhanced by the sigma factor (σ subunit) that associates with the RNAP core (Burgess & Anthony, 2001). The various types of sigma factors are largely classified into two classes, σ^{54} and σ^{70} , based on their structures. σ^{54} , also known as σ^N , often controls nitrogen metabolism, but other functions have been attributed in several organisms. Most sigma factors belong to σ^{70} , with four subgroups. Group I sigma factors are known as the primary sigma factors, which include *Escherichia coli* σ^{70} and *Thermus aquaticus* σ^A . The major function of the primary sigma factors is the transcriptional control of housekeeping genes (Kazmierczak *et al.*, 2005). Group I sigma factors are structurally divided into four highly conserved regions (regions 1–4), with further subdivision into two to four subregions. Region 1 regulates the DNA-binding affinity of the sigma factor by masking its DNA-binding domain (Dombroski *et al.*, 1992), whereas regions 2–4 directly interact with the DNA. Region 2 (specifically subregion 2.4) binds to the −10 promoter, whereas region 3 interacts upstream of the −10 promoter. Region 4 (specifically subregion 4.2) binds to the −35 promoter (Paget & Helmann, 2003). Group II sigma factors closely resemble group I sigma factors, but are not essential for cell growth. Group III sigma factors are sequentially distant from σ^{70} and usually activate regulons of specific signals such as those of the developmental checkpoint or heat shock. Group IV sigma factors, which form the largest and the most diverse subgroup, control the expression of genes that regulate many extracytoplasmic stimuli. Hence, the sigma factors in group IV are also called ECF (extracytoplasmic function) sigma factors. One interesting structural aspect of the ECF sigma factors is the absence of most or all of regions 1 and 3 (Paget *et al.*, 2002).

Our target protein, SigR (σ^R) from *Streptomyces coelicolor* A3(2), belongs to the ECF sigma factors and contains only the conserved regions 2 and 4. σ^R regulates oxidative stress in cooperation with the anti-sigma factor RsrA (regulator of SigR; Paget *et al.*, 1998). In a reducing environment, several cysteine thiols of RsrA in their reduced (−SH) form bind to zinc. In this state, RsrA sequesters σ^R to inhibit its ability to interact with RNAP. Oxidative stress leads the thiols of the RsrA to undergo disulfide-bond formation, and

consequently frees σ^R for association with RNAP, resulting in induction of the thioredoxin operon (*trxBA*) to relieve the stress (Kang *et al.*, 1999).

In order to understand the structural basis of σ^R interaction compared with other sigma factors, we have initiated overexpression, purification, crystallization and diffraction experiments of full-length *S. coelicolor* σ^R . Notably, only the structure of the amino-terminal region 2 of *S. coelicolor* σ^R has been reported (Li *et al.*, 2002). Diffracting crystals were obtained using hanging-drop vapour diffusion, and anomalous X-ray diffraction data were collected to 2.60 Å resolution using synchrotron radiation. Mass-spectrometric analysis of the harvested crystal as well as preliminary molecular-replacement results indicate that the asymmetric unit of the crystal contains only the carboxyl-terminal region 4, which is a proteolytic fragment of the full-length σ^R .

2. Materials and methods

2.1. Protein overexpression and purification

The full-length SigR (σ^R) from *S. coelicolor* (residues 1–227; gi:21223584) was PCR-amplified from a gene kindly provided by Dr Roe Jung-Hye (Seoul National University, Republic of Korea) and cloned into pET-28a vector (Novagen) using *Nde*I and *Bam*HI restriction sites. The correct sequence of the insert gene was verified by DNA sequencing using T7 promoter and terminator primers. The protein was overexpressed with an amino-terminal His₆ tag using plasmid-transformed *Escherichia coli* BL21 (DE3) (Novagen) cells. The cells were first grown at 37°C in minimal medium with kanamycin selection (25 µg ml^{−1}). At an *A*₆₀₀ of ~0.5, several amino acids such as lysine, phenylalanine, threonine, isoleucine, leucine and valine were added for inhibition of methionine biosynthesis. At an *A*₆₀₀ of ~0.6, expression of the selenomethionine-substituted recombinant protein was induced by 0.5 mM IPTG with the addition of seleno-

Table 1

Data-collection statistics.

Values in parentheses are for the outer shell.

Diffraction source	Beamline 5C, PLS
Wavelength (Å)	0.9792
Temperature (K)	100
Space group	<i>P</i> 4 ₃ 2 ₁ 2
Unit-cell parameters (Å)	<i>a</i> = <i>b</i> = 42.14, <i>c</i> = 102.02
Resolution range (Å)	50.0–2.60 (2.64–2.60)
No. of unique reflections	3168
Completeness (%)	99.9 (100)
Multiplicity	25.7 (26.7)
<i>I</i> / <i>σ</i> (<i>I</i>)	55.4 (8.8)
<i>R</i> _{merge} [†]	0.069 (0.481)
Wilson <i>B</i> factor (Å ²)	60.9‡
Data-processing software	HKL-2000

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl) \rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. [‡] Pearson correlation coefficient $R^2 = 0.9$.

methionine (25 mg per 500 ml of minimal medium). After induction, the cells were grown for 16 h at 22°C prior to harvesting by centrifugation at 4500g (10 min, 4°C). The cell pellet was resuspended in ice-cold lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole) and homogenized using sonication on ice. The cell lysate was centrifuged at 70 000g (30 min, 4°C). The supernatant containing the soluble protein was poured into an Ni–nitrilotriacetic acid (Ni–NTA) column (Qiagen) and washed with five column volumes of wash buffer (20 mM Tris pH 7.5, 20 mM imidazole, 500 mM NaCl). The protein was further eluted with elution buffer (20 mM Tris pH 7.5, 200 mM imidazole, 500 mM NaCl). The eluted fractions were checked for protein using a colorimetric assay (Bio-Rad) and were combined and treated with bovine thrombin (Invitrogen) for removal of the His₆ tag (16 h, 4°C). The protein was further applied onto a Superdex 200 HR26/60 sizing column connected to an ÄKTA FPLC system (GE Healthcare). The column had previously been equilibrated with gel-filtration buffer (50 mM Tris pH 7.5, 150 mM NaCl,

Figure 1

(a) Crystals of *S. coelicolor* σ^R and its SDS–PAGE analysis [inset; lane 1, expressed full-length σ^R (~25 kDa); lane M, protein marker (labelled in kDa); lane 3, harvested crystal; MS, region submitted to mass-spectrometric analysis]. (b) A representative diffraction image of the σ^R crystal.

Table 2

List of peptide fragments found from mass-spectrometric analysis of the harvested crystal.

Peptide sequence†	MH ⁺ (Da)	Hit No.	σ^R region
SAESQALDHLPSDVK	1711.81	33	4
IAVYLADVEGFAYK	1558.82	19	4
NPADAEDLVQETYAK	1663.78	7	2
DRGLVPAGAGESNEAK	1570.78	6	4
SAESQALDHLPSDVKQALQAIPEEFR	2994.49	2	4
QALQAIPEEFR	1301.69	1	4
EIADIMGTPIGTVMRS	1690.85	1	4

† Peptide hits counted for cross-correlation score (XCORR) > 3.5 (XCORR as defined by *Proteome Discoverer* 1.3; Thermo Scientific, USA).

5 mM DTT). The elution profile of the full-length protein showed a single major peak; the fractions comprising this peak were concentrated by centrifugation (Amicon). The final protein concentration ($\sim 30 \text{ mg ml}^{-1}$) was estimated from the A_{280} with a molar extinction coefficient ($21\,620 \text{ M}^{-1} \text{ cm}^{-1}$) calculated based on the absorbance of tryptophan and tyrosine residues (Gill & von Hippel, 1989).

2.2. Crystallization and X-ray data collection

Conditions for obtaining *S. coelicolor* σ^R crystals were screened using commercial solutions (Hampton Research) with the hanging-drop crystallization method in 24-well plates at 22°C. Crystallization drops were prepared by mixing 1 μl reservoir solution and 1 μl concentrated protein solution ($\sim 30 \text{ mg ml}^{-1}$). Tiny crystals appeared after 5 d and grew further to maximum size within 10 d. Crystals grew from two conditions consisting of either 0.2 M lithium sulfate, 0.1 M Tris pH 8.5, 20–30% (w/v) PEG 3350 or 0.2 M magnesium chloride, 0.1 M Tris pH 8.5, 20–30% (w/v) PEG 4000. Crystals were transferred to a cryoprotectant solution consisting of 15% glycerol added to the reservoir solution and were flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K using a CCD detector (ADSC Quantum 315r) on beamline 5C of the Pohang Light Source (PLS), Pohang, Republic of Korea. Data were processed using *DENZO* and *SCALEPACK* from the *HKL-2000* package (Otwinowski & Minor, 1997). The σ^R crystal packs into space group $P4_32_12$ (Table 1). Furthermore, several crystals were harvested and their contents were analyzed using either MALDI of the intact protein or LC-MS/MS of the tryptic digestion fragments.

3. Results and discussion

Recombinant *S. coelicolor* σ^R was overexpressed in *E. coli* in a soluble form and was purified with an overall yield of $\sim 10 \text{ mg}$ per litre of minimal medium culture. Crystals of the protein were grown in two crystallization conditions. The best diffracting crystal was obtained using a reservoir solution consisting of 0.2 M lithium sulfate, 0.1 M Tris pH 8.5, 20–30% (w/v) PEG 3350. Crystals grew to an optimal size for X-ray diffraction experiments within 10 d (Fig. 1). The approximate dimensions of the crystals were $20 \times 20 \times 50 \mu\text{m}$ (Fig. 1). The crystals diffracted to 2.60 Å resolution, and anomalous diffraction data were collected using synchrotron radiation (Fig. 1). A total of 3168 unique reflections were measured and merged with an R_{merge} of 6.9%. The merged data set was 99.9% complete overall (50–2.60 Å resolution). The crystal belonged to space group $P4_32_12$, with

unit-cell parameters $a = b = 42.14$, $c = 102.02$ Å. The statistics of the collected data are summarized in Table 1. According to calculation of the Matthews coefficient (Matthews, 1968), the crystal asymmetric unit containing the full-length protein gave an unlikely negative solvent content. Molecular replacement using *E. coli* σ_4^E (PDB entry 2h27; 26% identical in sequence to *S. coelicolor* σ_4^R ; Lane & Darst, 2006) as the search model gave a likely correct solution (McCoy *et al.*, 2005; *Phaser* statistics of RFZ = 5.1 and TFZ = 11.4) with only one molecule of C-terminal region 4 (σ_4^R) in the crystal asymmetric unit. Several crystals grown from the crystallization reservoir were harvested and gel-electrophoresed and their contents were analyzed using mass spectrometry (Fig. 1, inset and Table 2). Analysis of the trypsin-treated gel plug again suggested that the crystal contained mostly σ_4^R (Table 2). Based on the fragments identified by the MS/MS data, we believe that residues 134–223 of region 4 are contained in the crystal. Although a single peptide of σ^R region 2 was detected (Table 2), we conclude that the full-length σ^R underwent proteolysis into σ_4^R during the time necessary for crystal formation. The average intact mass of the crystal content measured by MALDI is 10.1 kDa (results not shown). The presence of one such molecule in the crystal asymmetric unit suggests a crystal volume per protein weight (V_M) of $2.24 \text{ Å}^3 \text{ Da}^{-1}$ and 45.1% solvent content. In the future, we plan to analyze the structure of *S. coelicolor* σ_4^R in comparison to other reported region 4 structures of sigma factors such as *E. coli* σ^E and *Mycobacterium tuberculosis* σ^L (or σ^C) (Lane & Darst, 2006; Thakur *et al.*, 2007, 2010).

We thank Professor Jung-Hye Roe (Seoul National University, Republic of Korea) for kindly providing the gene for *S. coelicolor* SigR. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST; 2013R1A1A2005276). Experiments at PLS were supported in part by MEST and Pohang University of Science and Technology (POSTECH).

References

- Burgess, R. R. & Anthony, L. (2001). *Curr. Opin. Microbiol.* **4**, 126–131.
- Dombroski, A. J., Walter, W. A., Record, M. T. Jr, Siegle, D. A. & Gross, C. A. (1992). *Cell*, **70**, 501–512.
- Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
- Kang, J.-G., Paget, M. S. B., Seok, Y.-J., Hahn, M.-Y., Bae, J.-B., Hahn, J.-S., Kleanthous, C., Buttner, M. J. & Roe, J.-H. (1999). *EMBO J.* **18**, 4292–4298.
- Kazmierczak, M. J., Wiedmann, M. & Boor, K. J. (2005). *Microbiol. Mol. Biol. Rev.* **69**, 527–543.
- Lane, W. J. & Darst, S. A. (2006). *PLoS Biol.* **4**, e269.
- Li, W., Stevenson, C. E. M., Burton, N., Jakimowicz, P., Paget, M. S. B., Buttner, M. J., Lawson, D. M. & Kleanthous, C. (2002). *J. Mol. Biol.* **323**, 225–236.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* **D61**, 458–464.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Paget, M. S. B. & Hermann, J. D. (2003). *Genome Biol.* **4**, 203.
- Paget, M. S. B., Hong, H.-J., Bibb, M. J. & Buttner, M. J. (2002). *Signals, Switches, Regulons, and Cascades: Control of Bacterial Gene Expression*, edited by D. A. Hodgson & C. M. Thomas, pp. 105–125. Cambridge University Press.
- Paget, M. S. B., Kang, J.-G., Roe, J.-H. & Buttner, M. J. (1998). *EMBO J.* **17**, 5776–5782.
- Thakur, K. G., Joshi, A. M. & Gopal, B. (2007). *J. Biol. Chem.* **282**, 4711–4718.
- Thakur, K. G., Praveena, T. & Gopal, B. (2010). *J. Mol. Biol.* **397**, 1199–1208.