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

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of hypothetical protein SCO4226 from *Streptomyces coelicolor* A3(2)

Shu Wang,[a,b] Yong-Xing He,[b] Rui Bao,[b] Yan-Bin Teng,[b] Bo-Ping Ye[a]* and Cong-Zhao Zhou[b]*

[a]School of Life Sciences and Technology, China Pharmaceutical University, Nanjing, Jiangsu 210009, People's Republic of China, and [b]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China

Correspondence e-mail: yebp2001@yahoo.com.cn, zcz@ustc.edu.cn

Received 14 July 2008
Accepted 10 August 2008

A non-Pfam hypothetical protein SCO4226 of molecular weight 9 kDa from *Streptomyces coelicolor* A3(2) was overexpressed in *Escherichia coli* and the purified recombinant protein was crystallized using the sitting-drop vapour-diffusion method. An X-ray diffraction data set was collected to 2.0 Å resolution. The crystal belonged to space group $P2_1$, with unit-cell parameters a = 29.67, b = 67.00, c = 34.43 Å, $\alpha = \gamma = 90.00$, $\beta = 94.26^\circ$.

1. Introduction

The *Streptomyces coelicolor* A3(2) open reading frame (ORF) *SCO4226* encodes an 82-residue protein of unknown function with a theoretical molecular weight of 9049 Da (Redenbach *et al.*, 1996). The results of searches using *BLASTP* and *PSI-BLAST* (http://www.ncbi.nlm.nih.gov; Altschul *et al.*, 1997) suggested that SCO4226 is similar in primary sequence to the N-termini of three proteins: a guanylate cyclase (YP_001978230; 40% identity), an AraC-family member (YP_860235; 39% identity) and a putative two-component system (PhoR/PhoP system) regulator protein (ZP_01106696; 35% identity) (Sola-Landa *et al.*, 2005; Fig. 1). To date, the structures of these proteins have not been reported.

Guanylate cyclases catalyze the cyclization of GTP and produce secondary messenger cGMP (Schulz *et al.*, 2008; Schmidt *et al.*, 2004). It has been proposed that the expression of soluble guanylate cyclase (sGC) might be alternated by increased superoxide production (Schulz *et al.*, 2008). Lineage analysis of sGC suggests that sGC was originally an O_2 sensor in primordial eukaryotes and then evolved to specifically recognize NO for more effective local cell-to-cell signalling (Fitzpatrick *et al.*, 2006).

Most AraC-family members are transcription regulators and share a DNA-binding helix–turn–helix motif within the homologous C-terminus (Gallegos *et al.*, 1993). However, the N-terminal regions of these regulators are highly variable and may be responsible for specific binding to different activator molecules (Gallegos *et al.*, 1993; Martin & Rosner, 2001). One member of the AraC family, *Escherichia coli* SoxS, the counterpart of which in *S. coelicolor* is located downstream of *SCO4226*, has been reported to be involved in the regulation of responses to oxidative stress (Wu & Weiss, 1991; Li & Demple, 1994).

The PhoR/PhoP system, which lies upstream of *SCO4226*, also regulates the expression of oxidative stress-related proteins (Ghorbel *et al.*, 2006; Sola-Landa *et al.*, 2008). The membrane-bound protein PhoR senses the phosphate-limitation signal and undergoes self-phosphorylation. Phosphorylated PhoR then transfers the phosphate group to PhoP, resulting in the subsequent expression of more than 30 genes including *Pho* regulons (Sola-Landa *et al.*, 2003). This system also plays an important role in the regulation of secondary metabolism and internal oxidative stress (Ghorbel *et al.*, 2006). The primary sequence homology between SCO4226, the AraC-family member and the PhoR/PhoP system regulator protein implies a possible relation of their functions, which needs to be proved by structure-based

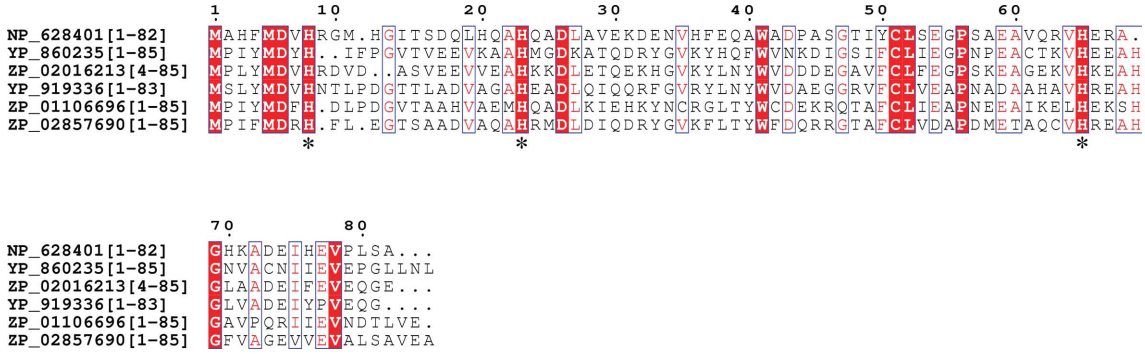


Figure 1
Sequence alignment of SCO4226 using the programs *MULTALIN* (Corpet, 1988) and *ESPript* (Gouet *et al.*, 2003) based on the results of *BLASTP* and *PSI-BLAST* searches (http://www.ncbi.nlm.nih.gov). NP_628401.1, SCO4226 from *S. coelicolor*; YP_860235.1, AraC-family transcriptional regulator protein from *Gramella forsetii*; ZP_01106696.1, putative two-component system sensor kinase from *Flavobacteriales bacterium*; ZP_02857690.1, adenylate/guanylate cyclase from *Rhizobium leguminosarum*; ZP_02016213.1, putative cyclase/kinase from *Halorubrum lacusprofundi*; YP_919336.1, putative guanylate cyclase protein from *Nocardioides* sp. JS614. Conserved residues are marked in red and conserved histidine residues are marked by asterisks.

functional interpretation and/or biochemical/biophysical investigations.

2. Materials and methods

2.1. Cloning and protein expression

The ORF encoding SCO4226 was amplified by PCR using the cosmid DNA of *S. coelicolor* A3(2) as a template (provided by Dr Marie-Joelle Virolle at the Institute of Genetics and Microbiology, University Paris-Sud XI). Primers for amplification were designed based on the sequence available from http://streptomyces.org.uk (*SCO4226*, 2SCD46.40c). *Nde*I and *Xho*I restriction sites were incorporated into the sequences of the sense and antisense primers 5′-CCACATATGGCTCACTTCATGGACGTACAC-3′ and 5′-ATTCTCGAGCTATGCCGACAGGGGG-3′, respectively. After digestion with *Nde*I and *Xho*I restriction enzymes, the PCR product was cloned into the expression vector pET22b. *E. coli* Top10 cells were transformed with the ligation mixture. The sequence of the construct, which had no tags or extra residues, was confirmed by DNA sequencing. *E. coli* Rosetta (DE3) cells were transformed with plasmid pET22b-*SCO4226* and grown on agar plates containing 50 µg ml^{-1} ampicillin. A single colony was verified and cultured in LB broth with 50 µg ml^{-1} ampicillin and 30 µg ml^{-1} chloramphenicol. Expression of the target protein was induced for 5 h at 310 K by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 m*M* when the OD_{600} reached 0.6. Cells were harvested by centrifugation at 7330*g* for 10 min and then resuspended in cold lysis buffer (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl). The suspension was sonicated and clarified by centrifugation at 29 300*g* for 25 min at 277 K.

2.2. Protein purification

The high histidine-residue content of SCO4226 enabled us to purify the protein by affinity chromatography. The supernatant was loaded onto a nickel-chelating Sepharose Fast Flow column (Amersham Biosciences) pre-equilibrated with buffer containing 20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl. After the column had been washed with 20 ml of the same buffer, the binding protein was eluted by a gradual increase in imidazole concentration from 10 to 200 m*M*. The fraction that was eluted by binding buffer containing 200 m*M* imidazole was then further purified on a HiLoad 16/60 Superdex 75 column (Amersham Biosciences) equilibrated with buffer containing 20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl. The fractions containing the target protein were verified by SDS–PAGE.

2.3. Crystallization and X-ray data collection

Crystals of SCO4226 were obtained by the sitting-drop vapour-diffusion method at 289 K using screening kits from Hampton Research. Each sitting drop consisted of 1.0 µl 10 mg ml^{-1} protein

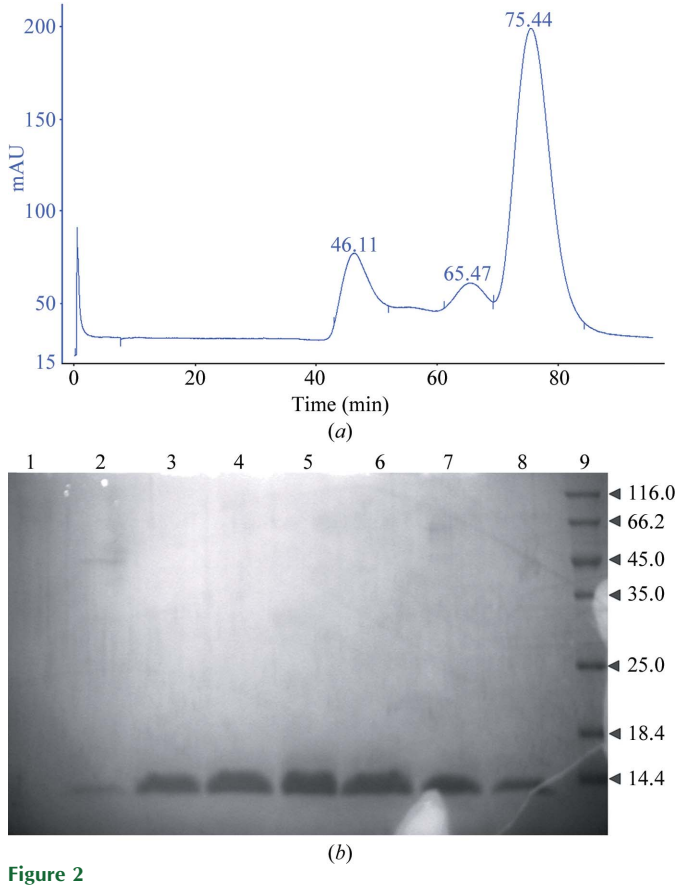


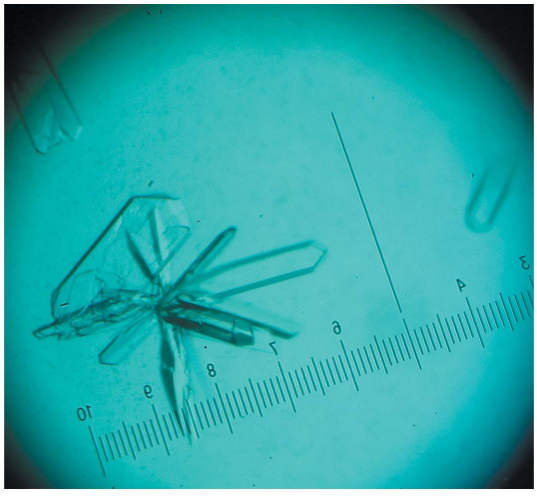
Figure 2
(*a*) Gel filtration of SCO4226 using a HiLoad 16/60 Superdex 75 column. The flow rate was 1 ml min^{-1}. Three main peaks were detected: peak I (46.11 min), peak II (65.47 min) and peak III (75.44 min). (*b*) 15% SDS–PAGE analysis of fractions from gel filtration. Lanes 1 and 2 correspond to the fractions in peaks I and II, respectively; lanes 3–8 correspond to fractions from peak III; lane 9 contains low-molecular-weight markers (kDa).

Table 1
Data-collection statistics.

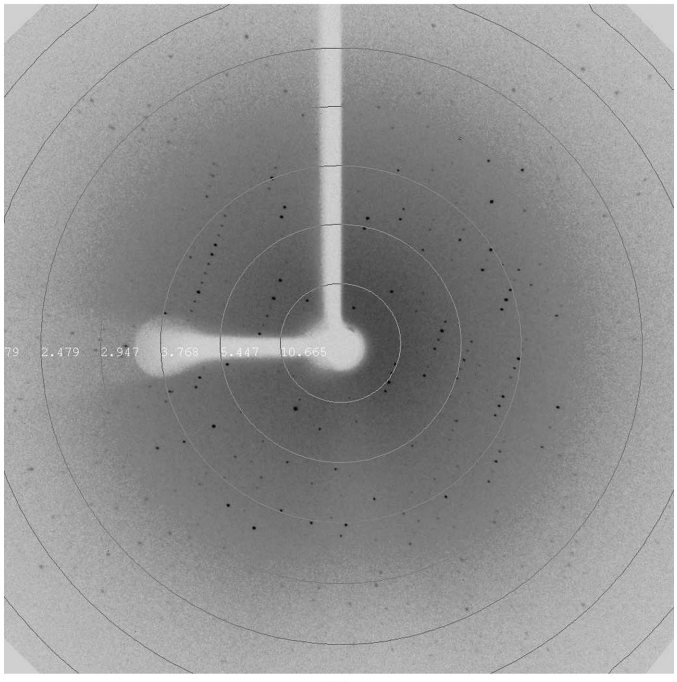
Values in parentheses are for the highest resolution shell.

Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 29.67, b = 67.00, c = 34.43, $\alpha = \gamma$ = 90.00, β = 94.26
Temperature (K)	100
Wavelength (Å)	1.54056
Resolution range (Å)	19.12–2.00 (2.11–2.00)
Total No. of observations	28975 (2570)
No. of unique reflections	9105 (1305)
Mean $I/\sigma(I)$	15.6 (3.7)
Completeness (%)	99.5 (98.3)
Redundancy	3.2 (2.0)
$R_{\rm merge}$ (%)	0.082 (0.196)
$R_{\rm p.i.m.}$†	0.053 (0.171)

† $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ are the observed intensities, $\langle I(hkl)\rangle$ are the average intensities and N is the multiplicity of reflection hkl. Calculated according to Weiss (2001) and Evans (2006).



(*a*)



(*b*)

Figure 3
(*a*) Crystals of SCO4226. (*b*) Diffraction pattern of a SCO4226 crystal.

solution (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl) and 1 µl reservoir solution and was equilibrated against 100 µl reservoir solution. Crystals suitable for X-ray diffraction appeared within 3 d using a reservoir consisting of 1.6 *M* sodium citrate dehydrate pH 6.5. Crystals were flash-frozen in liquid nitrogen using a cryoprotectant consisting of the reservoir solution plus 25% glycerol. Diffraction data were collected to 2.00 Å resolution at 100 K using a Gemini R Ultra mounted with a 135 mm diagonal Ruby CCD detector and co-mounted Enhance (Mo) and Enhance Ultra (Cu) X-ray sources (Guangzhou Institute of Biomedicine and Health, Chinese Academy of Sciences). The data were indexed and integrated with *CrysAlisPro* processing software (Oxford Diffraction) and scaled using *SCALA* (Evans, 1997) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Despite lacking a His tag, SCO4226 showed a moderate affinity for the nickel-chelating Sepharose column. We therefore speculated that it may possess potential metal-binding ability. Atomic absorption spectra indicated that the purified protein contains a traceable amount of nickel, with an occupancy of 11%. Inspection of the primary sequences revealed nine histidine residues, three of which are well conserved among different species (Fig. 1). It is likely that the function of SCO4226 involves metal-ion ligation and the conserved histidine residues might be clustered in the tertiary structure in order to coordinate a metal ion.

Gel filtration indicated that SCO4226 eluted at a volume corresponding to about 20 kDa, which is approximately twice its theoretical molecular weight of 9 049 Da (Fig. 2*a*). The purity was checked by SDS–PAGE (Fig. 2*b*). Liquid-chromatography mass spectrometry (LTQ, Thermo Fisher) was used to further ensure that the collected fractions contained the target protein. A detected peptide (AGHK-ADEIHEVPLSA) matched the last 15 residues of SCO4226 exactly, whereas no matching fragment was detected in the *E. coli* protein database.

Plate-like crystals appeared in a drop consisting of 1.6 *M* sodium citrate dehydrate pH 6.5 in 3 d (Fig. 3*a*). A set of diffraction data was collected from a single crystal to 2.00 Å resolution (Fig. 3*b*). The crystal belonged to space group $P2_1$, with unit-cell parameters $a = 29.67$, $b = 67.00$, $c = 34.43$ Å, $\alpha = \gamma = 90.00$, $\beta = 94.26^\circ$. Based on the Matthews coefficient calculation, it was possible that each asymmetric unit contained one molecule, with a Matthews coefficient of 3.77 Å^3 Da^{-1} and a solvent content of 67.39%, or two molecules, with a Matthews coefficient of 1.89 Å^3 Da^{-1} and a solvent content of 34.8% (Matthews, 1968). However, the observation of a twofold noncrystallographic axis might better support the notion of two molecules per asymmetric unit. The data-collection statistics are summarized in Table 1.

The gene cluster including *SCO4226* was proposed to be involved in regulation of phosphate metabolism and/or response to oxidative stress. An increasing number of studies of these genes and their encoded proteins have been reported (Ghorbel *et al.*, 2006; Sola-Landa *et al.*, 2005, 2008; Blanco *et al.*, 2002), but the function of SCO4226 still remains unknown. The structure-based functional interpretation will provide us with new information about this protein.

This research was supported by grants 30628011 and 30470366 from the Chinese National Natural Science Foundation and projects

2006CB910202 and 2006CB806501 from the Ministry of Science and Technology of China.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Blanco, A. G., Sola, M., Gomis-Rüth, F. X. & Coll, M. (2002). *Structure*, **10**, 701–713.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Corpet, F. (1988). *Nucleic Acids Res.* **16**, 10881–10890.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.

Fitzpatrick, D. A., O'Halloran, D. M. & Burnell, A. M. (2006). *BMC Evol. Biol.* **6**, 1–18.

Gallegos, M. T., Michán, C. & Ramos, J. L. (1993). *Nucleic Acids Res.* **21**, 807–810.

Ghorbel, S., Kormanec, J., Artus, A. & Virolle, M. J. (2006). *J. Bacteriol.* **188**, 677–686.

Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.

Li, Z. & Demple, B. (1994). *J. Biol. Chem.* **269**, 18371–18377.

Martin, R. G. & Rosner, J. L. (2001). *Curr. Opin. Microbiol.* **4**, 132–137.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H. & Hopwood, D. A. (1996). *Mol. Microbiol.* **21**, 77–96.

Schmidt, P. M., Schramm, M., Schröder, H., Wunder, F. & Stasch, J. P. (2004). *J. Biol. Chem.* **279**, 3025–3032.

Schulz, E., Jansen, T., Wenzel, P., Daiber, A. & Münzel, T. (2008). *Antioxid. Redox. Signal.* **10**, 1115–1126.

Sola-Landa, A., Moura, R. S. & Martín, J. F. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 6133–6138.

Sola-Landa, A., Rodríguez-García, A., Apel, A. K. & Martín, J. F. (2008). *Nucleic Acids Res.* **36**, 1358–1368.

Sola-Landa, A., Rodríguez-García, A., Franco-Domínguez, E. & Martín, J. F. (2005). *Mol. Microbiol.* **56**, 1373–1385.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

Wu, J. & Weiss, B. (1991). *J. Bacteriol.* **173**, 2864–2871.