Ryohei Nomoto,^a Takeaki
Tezuka,^a Ken-ichi Miyazono,^b
Masaru Tanokura,^b Sueharu
Horinouchi^{a‡} and Yasuo
Ohnishi^{a*}

^aDepartment of Biotechnology, Graduate School
of Agricultural and Life Sciences, The University
of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113-8657, Japan, and ^bDepartment of
Applied Biological Chemistry, Graduate School
of Agricultural and Life Sciences, The University
of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113-8657, Japan

‡ Deceased on 12 July 2009.

Correspondence e-mail:
ayasuo@mail.ecc.u-tokyo.ac.jp

Received 13 May 2012
Accepted 4 July 2012

© 2012 International Union of Crystallography
All rights reserved

Purification, crystallization and preliminary X-ray analysis of SGR6054, a *Streptomyces* homologue of the mycobacterial integration host factor mIHF

The mycobacterial integration host factor (mIHF) is a small nonspecific DNA-binding protein that is essential for the growth of *Mycobacterium smegmatis*. mIHF homologues are widely distributed among Actinobacteria, and a *Streptomyces* homologue of mIHF is involved in control of sporulation and antibiotic production in *S. coelicolor* A3(2). Despite their important biological functions, a structure of mIHF or its homologues has not been elucidated to date. Here, the *S. griseus* mIHF homologue (SGR6054) was expressed and purified from *Escherichia coli* and crystallized in the presence of a 16-mer duplex DNA by the sitting-drop vapour-diffusion method. The plate-shaped crystal belonged to space group C2, with unit-cell parameters $a = 88.53$, $b = 69.35$, $c = 77.71$ Å, $\beta = 96.63^\circ$, and diffracted X-rays to 2.22 Å resolution.

1. Introduction

Integration host factor (IHF) was originally discovered in *Escherichia coli* as an essential protein for the efficient site-specific integration and excision of bacteriophage λ DNA (Miller & Friedman, 1980). However, IHF is now recognized as a nucleoid-associated protein (NAP) in Gram-negative bacteria. It is involved in nucleoid structuring, chromosome replication, DNA rearrangement and transcription by binding to a well conserved nucleotide sequence and introducing a U-turn into the DNA (Dillon & Dorman, 2010). The mycobacterial integration host factor mIHF (105 amino acids) was also discovered as a protein that was able to stimulate the integrative recombination of mycobacteriophage L5 DNA (Pedulla *et al.*, 1996). Although mIHF is required for the formation of recombinogenic intasomes, which are nucleoprotein complexes that integrate bacteriophage DNA into the host DNA, it does not specifically bind to the integration site (L5 *attP*). Hence, mIHF has a nonspecific DNA-binding activity (Pedulla *et al.*, 1996; Lewis & Hatfull, 2003). Furthermore, mIHF has been shown to be essential for the growth of *Mycobacterium smegmatis*, suggesting that mIHF plays a pivotal role in the biology of mycobacteria (Pedulla & Hatfull, 1998). It should be noted that mIHF is not closely related to canonical IHF, the HU family of proteins or any other small DNA-binding proteins at the sequence level (Pedulla *et al.*, 1996; Pedulla & Hatfull, 1998). The quaternary structure of mIHF has not been elucidated. Homologues of mIHF are widely distributed among Actinobacteria, including mycobacteria, corynebacteria, streptomycetes and rhodococci.

The Gram-positive bacterial genus *Streptomyces* is characterized by its complex morphological differentiation culminating in spores and by its ability to produce a wide variety of secondary metabolites such as antibiotics, pigments, insecticides, immunosuppressants and antitumour agents (Flärdh & Buttner, 2009; van Wezel & McDowall, 2011). Very recently, the mIHF homologue (SCO1480; 107 amino acids) from a model *Streptomyces*, *S. coelicolor* A3(2), has been reported to have important functions in the control of sporulation and in antibiotic production (Yang *et al.*, 2012). In *SCO1480* null mutants sporulation was conditionally impaired and production of the pigmented antibiotics actinorhodin and undecylprodigiosin was enhanced. Similarly to mIHF, the recombinant SCO1480 protein

showed a nonspecific DNA-binding activity. Yang *et al.* (2012) suggested that the mIHF homologue SCO1480 acts as a global regulator of secondary metabolism and morphogenesis in *S. coelicolor* A3(2). mIHF homologues are highly conserved among *Streptomyces* species (more than 95% identity in amino-acid sequence). We have found that the mIHF homologue in *S. griseus* (SGR6054; NCBI reference sequence YP_001827566; 107 amino acids) is very important for the growth of this species (unpublished work). Despite their important biological functions, no structure of mIHF or of an mIHF homologue has been reported to date.

To elucidate the structural basis for nucleotide recognition by the mIHF homologue SGR6054, we attempted to cocrystallize SGR6054 with a 16-mer duplex DNA. Here, we report a crystal of SGR6054 (or an SGR6054–DNA complex) that diffracted X-rays to 2.22 Å resolution.

2. Materials and methods

2.1. Production and purification of the recombinant SGR6054 protein

Chromosomal DNA of *S. griseus* was used as a template for PCR amplification of the *SGR6054* coding sequence (Ohnishi *et al.*, 2008) using the following primers: 5'-CCGGAATTCATATGGCTCTTC-CGCCCTTACC-3' (*Eco*RI site underlined, *Nde*I site italicized and the start codon of *SGR6054* shown in bold) and 5'-GCCGGATCC-CTCGAGTCAGGCGGGGCTGCCGCCGAACT-3' (*Bam*HI site underlined, *Xho*I site italicized and the stop codon of *SGR6054* shown in bold). The PCR product was digested with *Eco*RI and *Bam*HI and cloned into pUC19 between the *Eco*RI and *Bam*HI sites. No PCR errors were found by nucleotide sequencing. The *Nde*I–*Xho*I fragment was excised from the pUC19 derivative and cloned into pET16b between the *Nde*I and *Xho*I sites (Novagen), resulting in pET16b-*SGR6054*. pET16b-*SGR6054* was introduced into *E. coli* BL21 (DE3) cells to produce N-terminally His-tagged SGR6054 with the sequence MGH₁₀SSGHIEGRH-*SGR6054*. Using factor Xa

treatment, the His tag was removed to yield recombinant SGR6054 protein with one additional His residue at the N-terminus.

The recombinant *E. coli* strain was grown at 310 K in 1.5 l Luria–Bertani (LB) broth containing 100 µg ml⁻¹ ampicillin. When the OD₆₀₀ reached 0.6–0.8, isopropyl β-D-1-thiogalactopyranoside (final concentration of 0.2 mM) was added and the culture was incubated for 16 h at 299.5 K. Cells were harvested by centrifugation at 5000g for 10 min at 277 K and resuspended in 20 ml buffer A (20 mM Tris–HCl, 500 mM NaCl, 10% glycerol pH 8.0) containing 10 mM imidazole. The cells were disrupted by sonication for 5 min. The crude cell-free extract was prepared by removal of cell debris by centrifugation at 10 000g for 30 min at 277 K followed by incubation with 1 ml Ni–NTA Superflow (Qiagen) which had been equilibrated with buffer A containing 10 mM imidazole for 30 min at 277 K with gentle agitation. The protein bound to the Ni–NTA Superflow was transferred into an empty column. The column was washed successively with buffer A containing 30 mM imidazole to remove contaminants. His-SGR6054 was then eluted with 10 ml buffer A containing 250 mM imidazole and dialyzed against buffer B (20 mM Tris–HCl, 100 mM NaCl, 10% glycerol pH 8.0). Subsequently, the N-terminal His tag was cleaved off with factor Xa for 12 h at 289 K. After the buffer had been exchanged to buffer C (20 mM MES, 10% glycerol pH 6.5) containing 50 mM NaCl by ultrafiltration (Millipore), the protein solution was applied onto a Toyopearl CM-650 cation-exchange column (Tosoh) which was equilibrated with buffer C containing 50 mM NaCl. Following washing with buffer C containing 200 mM NaCl, the His-tag-cleaved SGR6054 protein was eluted with buffer C containing 400 mM NaCl. After each purification step, the proteins were examined by SDS–PAGE. The His-tag-cleaved SGR6054 protein was concentrated to 3 mg ml⁻¹ by ultrafiltration in buffer C containing 100 mM NaCl. The protein concentration was measured using the Bradford reagent (Bio-Rad).

2.2. Crystallization

For cocrystallization of the protein–DNA complex, a 16-mer duplex DNA (5'-AATCGAACCCGCGTAG-3') which was designed based on the *attB* site of *S. griseus* was mixed with equimolar SGR6054 and concentrated to 200 µM. All crystallization experiments were performed at 293 K by the sitting-drop vapour-diffusion method. In the initial screening, droplets consisting of 0.75 µl protein–DNA solution and 0.75 µl reservoir solution were equi-

Figure 1

Purification of the recombinant SGR6054 protein. Proteins were examined by 15% SDS–PAGE. Lane M, protein molecular-weight markers (labelled in kDa); lane 1, supernatant after sonication of the recombinant *E. coli* cells; lane 2, pooled fractions from Ni–NTA Superflow; lane 3, after factor Xa treatment; lanes 4–6, eluted fractions containing SGR6054 from cation-exchange chromatography.

Figure 2

Crystals of SGR6054 obtained by the sitting-drop vapour-diffusion method.

brated against 50 μ l reservoir solution. An initial crystallization trial was carried out using the commercial crystal screening kits Crystal Screen HT and Index HT (Hampton Research) and the Nucleix Suite (Qiagen). Crystals appeared in several days using a condition from the Nucleix Suite (0.5 mM spermine, 20 mM MgCl_2 , 50 mM sodium succinate pH 5.5, 3.0 M ammonium sulfate). Crystallization drops were then prepared by mixing 1.5 μ l protein–DNA solution and 1.5 μ l reservoir solution in order to obtain larger crystals.

2.3. Data collection and processing

The crystals were mounted on cryoloops and flash-cooled at 95 K in a nitrogen stream for data collection. For cryoprotection, the SGR6054 crystals were soaked in reservoir solution supplemented with 20% (v/v) glycerol for a few seconds. An X-ray diffraction data set was collected on the NW12A beamline at the Photon Factory Advanced Ring (PF-AR), Tsukuba, Japan using an ADSC Quantum 210 CCD detector and an X-ray wavelength of 1.0000 Å. A total of 360 images were collected with an oscillation range of 0.5° and a crystal-to-detector distance of 212 mm. The diffraction data were indexed, integrated and scaled with *XDS* (Kabsch, 2010).

Figure 3
X-ray diffraction image from an SGR6054 crystal. The circle indicates a resolution of 2.20 Å.

Table 1

Diffraction data statistics for the SGR6054 crystal.

Values in parentheses are for the highest resolution shell.

Beamline	PF-AR NW12A
Wavelength (Å)	1.0000
Space group	C2
Unit-cell parameters (Å, °)	$a = 88.53$, $b = 69.35$, $c = 77.71$, $\beta = 96.63$
Resolution range (Å)	20.0–2.22 (2.28–2.22)
Multiplicity	4.86 (4.62)
Completeness (%)	97.3 (96.8)
$R_{\text{merge}}^{\dagger}$ (%)	7.7 (45.3)
Average $I/\sigma(I)$	16.29 (4.43)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity over all equivalent reflections.

3. Results and discussion

The recombinant SGR6054 protein was produced in a soluble form in *E. coli* and purified by two column-chromatography steps. About 3 mg SGR6054 was purified from 1.5 l culture. The purity of SGR6054 was greater than 95% as determined by SDS–PAGE (Fig. 1). Using crystallization screening, diffraction-quality crystals were obtained by the sitting-drop vapour-diffusion method with a reservoir solution composed of 0.5 mM spermine, 20 mM MgCl_2 , 50 mM sodium succinate pH 5.5, 3.0 M ammonium sulfate. An image of a typical SGR6054 crystal (0.1 $\times$ 0.07 $\times$ 0.02 mm) is shown in Fig. 2. A single crystal was carefully separated from a cluster like that shown in Fig. 2 using a small needle. A diffraction data set was processed to 2.22 Å resolution (Fig. 3).

The crystal belonged to the *C*-centred monoclinic space group C2, with unit-cell parameters $a = 88.53$, $b = 69.35$, $c = 77.71$ Å, $\beta = 96.63^\circ$. Two or three SGR6054–DNA complex molecules were predicted to be present in the asymmetric unit, corresponding to a Matthews coefficient of 2.80 Å³ Da^{−1} and a solvent content of 56.14% or to a Matthews coefficient of 1.87 Å³ Da^{−1} and a solvent content of 34.22%, respectively (Matthews, 1968). Although the recombinant SGR6054 protein was crystallized in the presence of a 16-mer duplex DNA, it is possible that the crystal did not contain a DNA molecule. Because the crystal was formed in the presence of 3 M ammonium sulfate, an SGR6054–DNA complex seemed to be rather unlikely. In this case, four SGR6054 molecules were predicted to be present in the asymmetric unit, corresponding to a Matthews coefficient of 2.57 Å³ Da^{−1} and a solvent content of 52.13%. To determine the location of the noncrystallographic symmetry, a self-rotation function was calculated using *MOLREP* (Vagin & Teplyakov, 2010) in the resolution range 20.0–3.5 Å. However, the result was inconclusive and therefore the noncrystallographic symmetry will be determined during subsequent structure-solution steps. Data-collection statistics are summarized in Table 1. The C-terminal region of SGR6054 shares 36% sequence identity with the N-terminal region of bacterial 30S ribosomal protein S13. We attempted to use the S13 protein (PDB entry 1fjg, residues 14–59 of chain M; Carter *et al.*, 2000) as a model to solve the structure of SGR6054 by the molecular-replacement method, but failed to obtain any satisfactory results. We have recently succeeded in obtaining crystals of selenomethionine-containing SGR6054 using a similar method and we plan to solve the crystal structure using the multiple-wavelength anomalous dispersion (MAD) method.

The synchrotron-radiation experiments were performed on beamline NW-12A at the Photon Factory (Proposal No. 2008S2-001). This work was supported in part by the Targeted Proteins Research

Program (TPRP) of the Ministry of Education, Culture, Sports, Science and Technology, Japan and a Funding Program for Next Generation World-Leading Researchers from the Bureau of Science, Technology and Innovation Policy, Cabinet Office, Government of Japan.

References

- Carter, A. P., Clemons, W. M. Jr, Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 340–348.
- Dillon, S. C. & Dorman, C. J. (2010). *Nature Rev. Microbiol.* **8**, 185–195.
- Flärdh, K. & Buttner, M. J. (2009). *Nature Rev. Microbiol.* **7**, 36–49.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Lewis, J. A. & Hatfull, G. F. (2003). *J. Mol. Biol.* **326**, 805–821.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Miller, H. I. & Friedman, D. I. (1980). *Cell*, **20**, 711–719.
- Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. & Horinouchi, S. (2008). *J. Bacteriol.* **190**, 4050–4060.
- Pedulla, M. L. & Hatfull, G. F. (1998). *J. Bacteriol.* **180**, 5473–5477.
- Pedulla, M. L., Lee, M. H., Lever, D. C. & Hatfull, G. F. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 15411–15416.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- van Wezel, G. P. & McDowall, K. J. (2011). *Nat. Prod. Rep.* **28**, 1311–1333.
- Yang, Y.-H., Song, E., Willemse, J., Park, S.-H., Kim, W.-S., Kim, E.-J., Lee, B.-R., Kim, J.-N., van Wezel, G. P. & Kim, B.-G. (2012). *Antonie Van Leeuwenhoek*, **101**, 479–492.