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

Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH from *Escherichia coli* and its complex with ops DNA

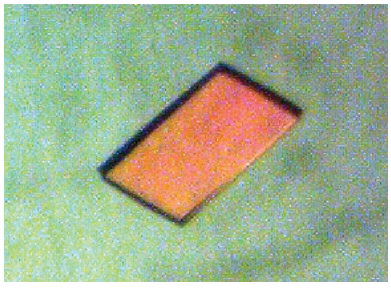
Marina N. Vassylyeva,[a]‡ Vladimir Svetlov,[b]‡ Sergiy Klyuyev,[a] Yancho D. Devedjiev,[a] Irina Artsimovitch[b]* and Dmitry G. Vassylyev[a]*

[a]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 402B Kaul Genetics Building, 720 20th Street South, Birmingham, AL 35294, USA, and [b]Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, USA

‡ These authors made an equal contribution.

Correspondence e-mail: artsimovitch.1@osu.edu, dmitry@uab.edu

Received 10 August 2006
Accepted 9 September 2006



The bacterial transcriptional factor and virulence regulator RfaH binds to rapidly moving transcription elongation complexes through specific interactions with the exposed segment of the non-template DNA strand. To elucidate this unusual mechanism of recruitment, determination of the three-dimensional structure of RfaH and its complex with DNA was initiated. To this end, the *Escherichia coli rfaH* gene was cloned and expressed. The purified protein was crystallized by the sitting-drop vapor-diffusion technique. The space group was $P6_122$ or $P6_522$, with unit-cell parameters $a = b = 45.46$, $c = 599.93$ Å. A complex of RfaH and a nine-nucleotide oligodeoxyribonucleotide was crystallized by the same technique, but under different crystallization conditions, yielding crystals that belonged to space group $P1$ (unit-cell parameters $a = 36.79$, $b = 44.01$, $c = 62.37$ Å, $\alpha = 80.62$, $\beta = 75.37$, $\gamma = 75.41^\circ$). Complete diffraction data sets were collected for RfaH and its complex with DNA at 2.4 and 1.6 Å resolution, respectively. Crystals of selenomethionine-labeled proteins in both crystal forms were obtained by cross-microseeding using the native microcrystals. The structure determination of RfaH and its complex with DNA is in progress.

1. Introduction

RfaH is a bacterial transcriptional factor that regulates the production of the LPS core and capsules and the expression of secreted toxins such as hemolysin (Bailey *et al.*, 1997). RfaH is a fast-evolving paralog of the essential bacterial transcription factor NusG, but their sequence homology is limited (the *Escherichia coli* RfaH and NusG proteins are only 17% identical and this identity is uniformly distributed over the entire protein length). Thus, although the *Aquifex aeolicus* NusG structure is available (Steiner *et al.*, 2002), this low sequence homology does not allow us to predict essential similarity at the structural level. NusG proteins from some species have been shown to bind to RNA and DNA non-specifically and cooperatively. However, they do not possess any known sequence specificity (Steiner *et al.*, 2002). In contrast, RfaH function depends on a 12-nucleotide (nt) regulatory site called the **o**peron **p**olarity **s**uppressor (ops), which is typically located within the 5′-proximal transcribed sequence of operons under RfaH control. The ops sites exhibit very high sequence conservation and similar location within homologous operons in different bacteria (Carter *et al.*, 2004) and are sufficient for RfaH recruitment to the transcription complexes *in vitro* (Artsimovitch & Landick, 2002). In contrast, RfaH orthologs display significant sequence divergence and, while interacting with the ops element similarly, are apparently optimized for interactions with their cognate RNA polymerases (RNAPs; Carter *et al.*, 2004).

When recruited to transcription elongation complexes (TEC), RfaH reduces pausing and termination by RNAP, thereby allowing efficient transcription of the distal genes within long multi-cistronic operons. RNA release at terminators is preceded by pausing, which is caused in turn by conformational changes in the TEC compromising productive alignment of nucleotides in the active site of RNAP (Artsimovitch & Landick, 2000; Toulokhonov *et al.*, 2001). Termination and pausing are modulated by various bacterial transcriptional regulators that act as enhancers or inhibitors of elongation, NusA and NusG (Burns *et al.*, 1998), accessory proteins, NusB and NusE

(Mason *et al.*, 1992), or antiterminators, λ N and Q (Rees *et al.*, 1997; Roberts *et al.*, 1998). Antitermination of transcription is typically mediated by multi-protein assemblies (Torres *et al.*, 2004; Weisberg & Gottesman, 1999); in contrast, RfaH increased the rate of transcription in the absence of other regulatory proteins (Artsimovitch & Landick, 2002). Another important aspect of the RfaH mechanism is the use of a specific sequence on the exposed non-template DNA strand for recruitment to the TEC. While this mode of recruitment is reminiscent of the bacterial σ-factors binding to the promoter DNA (Roberts & Roberts, 1996), binding of σ to RNAP precedes promoter recognition and commencement of catalysis, whereas RfaH recognizes the ops DNA in transcribing the TEC and does not bind strongly to RNAP alone (Artsimovitch & Landick, 2002).

RfaH is the only known transcriptional regulator that uses sequence-specific interactions with the non-template DNA strand to induce conformational changes that increase the processivity of the TEC. To elucidate the mechanism of action of RfaH, we sought to determine its three-dimensional structure as well as the mode of interaction with the target DNA. To this end, we have cloned and expressed the *rfaH* gene from *E. coli*. The purified protein and the complex with DNA have been crystallized and complete data sets have been collected.

2. Experimental procedures and results

2.1. Cloning, expression and purification

The RfaH expression vector pVS12 was constructed by subcloning the *E. coli rfaH* open reading frame (orf) into the *Nde*I and *Sap*I sites of the pTYB1 expression vector (New England Biolabs). To this end, the *rfaH* orf was amplified using *Pfu* DNA polymerase-driven high-fidelity PCR from plasmid pIA238 (Artsimovitch & Landick, 2002) and primers IA58 (GACGGTCATATGCAATCCTGGTATTTAC) and IA278 (GCAGTGGCTCTTCCGCACGCTTTGCGGAACTC-GGTGTTCTTCACA). The sequence of the construct was verified by automated sequencing (Plant-Microbe Genomics Facility, Ohio State University). In the constructed plasmid, RfaH is expressed under control of the T7 gene 10 promoter and is fused at its C-terminus to a chitin-binding protein/intein moiety. Following intein-mediated cleavage, the recombinant RfaH protein carries a C-terminal Ala residue instead of the genome-encoded Leu; this replacement was made to increase efficiency of cleavage and did not impact on the activity of the protein (data not shown). Overexpression was induced according to the Overnight Express autoinduction system (Novagen) in *E. coli* strain BL21 (λDE3) transformed with pVS12.

For purification of the RfaH protein (162 amino acids, 18 kDa), cells were harvested and disrupted by ultrasonication in ImpactCN500 buffer (50 m*M* Tris–HCl pH 8.8, 500 m*M* NaCl, 1 m*M* EDTA, 0.1 m*M* PMSF) with the addition of EDTA-free Complete Protease Inhibitor cocktail (Roche) and 0.1% Tween-20. The cell extract was clarified by centrifugation (27 000*g*, 15 min at 277 K) and loaded onto a disposable gravity column with chemically cross-linked chitin polymer (New England Biolabs) equilibrated with the same buffer. The column was then washed with ten volumes of ImpactCN500 buffer. To elicit the intein-mediated cleavage reaction, the column was finally equilibrated with three volumes of ImpactCN500 buffer with the addition of 50 m*M* 2-mercaptoethanol (ME) and incubated overnight at room temperature. The cleaved protein was eluted with three column volumes of HepA buffer (50 m*M* Tris–HCl pH 6.9, 5% glycerol, 1 m*M* ME, 0.1 m*M* PMSF). Fractions containing the protein of interest were combined and dialyzed against HepA buffer. After dialysis, the protein was loaded onto a Heparin FF 16/10 column (GE Healthcare) at 2 ml min^{-1} using an AKTA Purifier system. The column was washed with five volumes of HepA buffer followed by a 400 ml NaCl gradient from 0 to 600 m*M*. Fractions containing RfaH protein were combined and the purified protein was concentrated to approximately 3 mg ml^{-1} using an Amicon Ultra-15 10 000 Da molecular-weight cutoff concentrator and loaded onto HiLoad Superdex 75 16/60 (GE Healthcare) equilibrated with 750 m*M* NaCl in HepA buffer using an AKTA Purifier system at 0.5 ml min^{-1}. The purified protein was eluted with the same buffer and remained stable at 277 K without degradation for several weeks. The yield was 14 mg of protein per litre of culture.

To prepare the SeMet-labeled variant of RfaH, the plasmid pVS12 was introduced into the methionine-auxotrophic *E. coli* expression

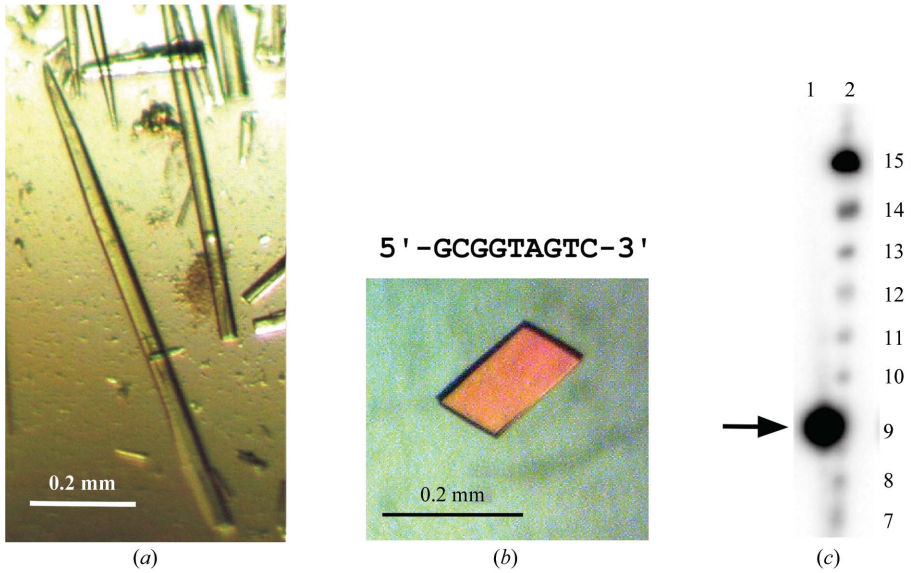


Figure 1
Crystallization of RfaH and its DNA complex. (*a*) Crystals of RfaH protein and (*b*) of the RfaH–ops complex. The sequence of the 9 nt oligonucleotide used for complex assembly is indicated in (*b*). (*c*) Gel electrophoretic analysis of the crystalline RfaH–ops sample. Lane 1 was loaded with the dissolved crystals and lane 2 with the oligonucleotide markers of indicated length. The labeled DNA oligonucleotide from the crystal is indicated by an arrow (see text for details).

Table 1
Data-collection statistics for RfaH and the RfaH–ops complex.

Values in parentheses are for the highest resolution shell.

	RfaH	RfaH–ops complex
Space group	$P6_122$ or $P6_522$	$P1$
Unit-cell parameters		
a (Å)	45.46	36.79
b (Å)	45.46	44.01
c (Å)	599.93	62.37
α (°)	90	80.62
β (°)	90	75.37
γ (°)	120	75.41
Source	22ID, SER-CAT, APS	22ID, SER-CAT, APS
Wavelength (Å)	1.0	1.0
Temperature (K)	100	100
Molecules in ASU	1–2	1–2
Solvent content (%)	75–50	73–45
Resolution (Å)	40–2.4 (2.49–2.4)	50–1.6 (1.66–1.6)
No. of observations	62701	117537
No. of unique reflections	15331	43553
Multiplicity	4.1 (2.9)	2.7 (2.1)
R_{merge}†	0.071 (0.373)	0.065 (0.497)
Completeness (%)	95.1 (89.9)	90.7 (87.9)
$I/\sigma(I)$	16.7 (2.0)	20.5 (2.1)

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

strain B834(DE3) (Novagen). Production of SeMet-labeled protein was carried out according to Studier's autoinduction protocol (Studier, 2005). SeMet-labeled RfaH was purified according to the protocol described above.

2.2. Crystallization of RfaH

Prior to crystallization, RfaH protein was dialyzed against 10 m*M* Tris–HCl buffer pH 7.8 with the addition of 50 m*M* KCl and 2 m*M* dithiothreitol (crystallization buffer) and was concentrated to approximately 16 mg ml^{-1}. The Hampton Crystal Screen kit (Jancarik & Kim, 1991) was used to determine initial crystallization conditions for the RfaH protein. Crystallization was carried out by the sitting-drop vapor-diffusion technique at 293 and at 277 K. Drops containing 1.5 µl precipitant solution and 1.5 µl protein solution were equilibrated against 0.5 ml reservoir solution. After 2 d of equilibration at 277 K, needle-like crystals grew in Crystal Screen II solution No. 26 [30%(*w*/*v*) PEG monomethyl ether 5000, 100 m*M* MES pH 6.5, 200 m*M* ammonium sulfate]. The initial conditions were optimized using Additive Screen I solution No. 4 (10 m*M* $CoCl_2$), which improved the size and the quality of the crystals. The crystals were then subjected to macroseeding using drops prepared under the same conditions. After a few days of equilibration, the crystals grew to maximum dimensions of 0.7 × 0.05 × 0.05 mm (Fig. 1*a*) and diffracted to beyond 3.0 Å resolution using an in-house X-ray source. They belonged to space group $P6_122$ or $P6_522$, with unit-cell parameters $a = b = 45.46$, $c = 599.93$ Å. The best crystals were obtained by macroseeding of drops prepared by mixing protein solution (1.5 µl) at 16 mg ml^{-1} with final optimized reservoir solution (1.5 µl) containing 20% PEG monomethyl ether 5000, 50 m*M* sodium cacodylate pH 6.5 and with the addition of 0.3 µl 100 m*M* $CoCl_2$.

2.3. Preparation and crystallization of RfaH–ops complex

The RfaH recognition sequence, ops (GGCGGTAGTCTG), is sufficient to mediate RfaH binding to the TEC (Artsimovitch & Landick, 2002). Because only a subset of the nucleotides in the ops motif (~9–10) would be exposed on the surface of the TEC for RfaH binding, variable lengths of the targeted non-template DNA were used for complex assembly. Oligonucleotides used for cocrystallization with RfaH contained 8–11 nt segments of the ops element, 5′-AGCGGTAGTCT (1a), 5′-GGCGGTAGTCT (1b), 5′-GGCGGTAGTC (2), 5′-GCGGTAGTC (3) and 5′-GGCGGTAG (4). The HPLC-purified oligonucleotides were purchased from Integrated DNA Technologies. Stock solutions were prepared in water at a concentration of 10 m*M* and were used without further purification.

The complex of RfaH protein with a DNA oligomer was prepared at room temperature. Aliquots of a 10 m*M* stock solution of one of the four DNA fragments above were diluted with crystallization buffer containing an additional 10 m*M* $MgCl_2$ and then mixed with an equal volume of RfaH (16 mg ml^{-1}) to form an equimolar ratio. The final concentration of the protein and DNA in the mixture was 0.44 m*M* and the Mg^{2+} concentration was 4.5 m*M*.

Hampton Crystal Screens (Jancarik & Kim, 1991) were used to determine the initial crystallization conditions for the complexes of RfaH with DNA fragments. Crystallization was carried out by the sitting-drop vapor-diffusion technique at 277 K. The Crystal Screen precipitant solutions were diluted twofold with water in the initial screening for RfaH–ops complexes in order to avoid a large number of immediately precipitated drops. Drops containing 1 µl diluted precipitant solutions and 1 µl complex solution were equilibrated against 0.5 ml of the same diluted precipitant solutions. After 2 d of equilibration at 277 K, crystals were obtained in Natrix Screen solution No. 31 [12.5%(*v*/*v*) PEG 550 monomethyl ether, 25 m*M* HEPES pH 7.0, 2.5 m*M* $MgCl_2$] using fragments 1b, 2 and 3. The best diffracting crystals grew when the 9 nt oligonucleotide (fragment 3, see above) was used for complex assembly with RfaH. These crystals were subjected to macroseeding using drops prepared under the same conditions. After a few days of equilibration, the crystals grew to typical dimensions of 0.2 × 0.1 × 0.02 mm (Fig. 1*b*) and diffracted to beyond 3.0 Å resolution using an in-house X-ray generator. They belonged to space group $P1$, with unit-cell parameters $a = 36.79$, $b = 44.01$, $c = 62.37$ Å, $\alpha = 80.62$, $\beta = 75.37$, $\gamma = 75.41°$. The final optimized reservoir solution used to grow the $P1$ crystals contained 22.2% PEG monomethyl ether 500, 44.4 m*M* Na HEPES pH 7.0, 4.4 m*M* $MgCl_2$.

Ten large crystals from RfaH–ops trials were harvested and were washed for ~1.5 h in 100 µl drops containing mother-liquor solution without oligonucleotides. In the course of this procedure, the crystals were transferred to new mother-liquor drops five times to ensure removal of the oligonucleotides sticking to the crystal surface, as well as those unbound nucleotides that might be present in the solvent region inside the crystals. The washed crystals were dissolved in 20 m*M* Tris–HCl pH 8.0, 20 m*M* NaCl, 5 m*M* $MgCl_2$, 0.1 m*M* DTT and labeled with (γ^{32}P)-ATP (GE Healthcare) and T4 polynucleotide kinase (New England Biolabs). Samples were analyzed by electrophoresis in a denaturing 10% acrylamide (19:1) gel (7 *M* urea, 0.5× TBE) along with (γ^{32}P)-labeled oligonucleotides of known length and concentration (Fig. 1*c*). This analysis confirmed that the 9-mer ops was specifically incorporated into the crystal lattice.

2.4. Data collection

Diffraction data were collected at SER-CAT beamline 22ID (APS, Argonne) under cryogenic conditions (Table 1). The RfaH crystals were flash-frozen using the mother-liquor solution with the addition of 25%(*v*/*v*) glycerol as a cryoprotectant; the crystals were soaked in this solution for a few seconds. RfaH–ops crystals were very unstable in all cryoprotectant solutions tested and for this reason we submerged the crystals in paraffin oil before flash-freezing. The data were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). To phase the diffraction data, we

prepared RfaH protein containing selenomethionine (SeMet) for crystallization. SeMet RfaH and SeMet RfaH–ops crystals were grown under the same conditions as the native proteins using seeds prepared from native crystals. The native crystals were crushed with a needle (Hampton Research), which was then plunged into fresh crystallization drops containing SeMet protein. Since the SeMet-labeled crystals are of the same diffraction quality as those of the native protein, we will undertake the structure solution of RfaH and its complex with DNA by multi-wavelength anomalous dispersion (MAD) methods using the Se atoms incorporated in the three Met residues (including the N-terminal Met residue) of each molecule.

We thank Dr A. Perederina for assistance in crystallization and data collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research under contract No. W-31-109-Eng-38. This work was supported in part by grants GM74252 (to DGV) and GM67153 (to IA) from the National Institutes of Health.

References

Artsimovitch, I. & Landick, R. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 7090–7095.
Artsimovitch, I. & Landick, R. (2002). *Cell*, **109**, 193–203.
Bailey, M. J., Hughes, C. & Koronakis, V. (1997). *Mol. Microbiol.* **26**, 845–851.
Burns, C. M., Richardson, L. V. & Richardson, J. P. (1998). *J. Mol. Biol.* **278**, 307–316.
Carter, H. D., Svetlov, V. & Artsimovitch, I. (2004). *J. Bacteriol.* **186**, 2829–2840.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Mason, S. W., Li, J. & Greenblatt, J. (1992). *J. Biol. Chem.* **267**, 19418–19426.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rees, W. A., Weitzel, S. E., Das, A. & von Hippel, P. H. (1997). *J. Mol. Biol.* **273**, 797–813.
Roberts, C. W. & Roberts, J. W. (1996). *Cell*, **86**, 495–501.
Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H. & Roberts, C. W. (1998). *Cold Spring Harb. Symp. Quant. Biol.* **63**, 319–325.
Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R. & Wahl, M. C. (2002). *EMBO J.* **21**, 4641–4653.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Torres, M., Balada, J. M., Zellars, M., Squires, C. & Squires, C. L. (2004). *J. Bacteriol.* **186**, 1304–1310.
Toulokhonov, I., Artsimovitch, I. & Landick, R. (2001). *Science*, **292**, 730–733.
Weisberg, R. A. & Gottesman, M. E. (1999). *J. Bacteriol.* **181**, 359–367.