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

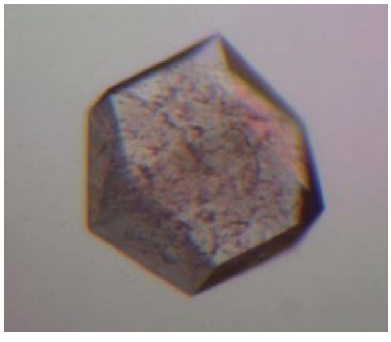
Crystallization and preliminary X-ray analysis of a cohesin-like module from AF2375 of the archaeon *Archaeoglobus fulgidus*

Milana Voronov-Goldman,[a,b] Ilit Noach,[c] Raphael Lamed,[a,b] Linda J. W. Shimon,[d] Ilya Borovok,[a] Edward A. Bayer[c]* and Felix Frolow[a,b]*

[a]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel, [b]The Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel, [c]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel, and [d]Department of Chemical Support, The Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: ed.bayer@weizmann.ac.il, mbfrolow@post.tau.ac.il

Received 28 December 2008
Accepted 23 January 2009



A cohesin-like module of 160 amino-acid residues from the hypothetical protein AF2375 of the noncellulolytic, hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* was cloned, expressed, purified, crystallized and subjected to X-ray structural study in order to compare its structure with those of cellulolytic cohesins. The crystals had cubic symmetry, with unit-cell parameters $a = b = c = 101.75$ Å in space group $P4_332$, and diffracted to 1.82 Å resolution. The asymmetric unit contained a single cohesin molecule. A model assembled from six cohesin structures (PDB entries 1anu, 1aoh, 1g1k, 1qzn, 1zv9 and 1tyj) of very low sequence identity to the cohesin-like module was used in molecular-replacement attempts, producing a marginal solution.

1. Introduction

Cohesins and dockerins are modular components that were initially discovered and characterized in cellulosomes (Salamitou *et al.*, 1992; Bayer *et al.*, 1994; Yaron *et al.*, 1995). These complementary modules participate in very high-affinity protein–protein interactions through which the dockerin-containing enzymatic subunits of the cellulosome are incorporated into the cohesin-bearing scaffoldin framework, thereby forming the multi-enzyme complex.

The presence of a cohesin or dockerin in a given protein had been considered to be a signature of the cellulosome (Bayer *et al.*, 1998). It was therefore surprising to discover cohesin-like and dockerin-like sequences in the genome of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*, a microbe that lacks glycoside hydrolases and thus clearly lacks a cellulosome (Bayer *et al.*, 1999). The *A. fulgidus* genome encodes two hypothetical proteins that include such modules. One protein (AF2375) contains a putative cohesin and an adjacent dockerin and the second protein (AF2376) contains a single cohesin. Using a cohesin–dockerin microarray approach, it was revealed that the AF2375 dockerin interacts with both its own cohesin and the cohesin of the tandem gene AF2376 (Haimovitz *et al.*, 2008). The interaction appears to be species-specific, since the *A. fulgidus* cohesin and dockerin failed to interact with known cohesins and dockerins from other species.

To date, the crystal structures of cohesins have mainly been determined from cellulolytic bacteria. More recently, the structure of a cohesin module from a glycoside hydrolase of the myonecrotic human pathogen *Clostridium perfringens* has been determined and was found to resemble known cohesin structures (Chitayat *et al.*, 2008) and to have the distinctive elongated nine-stranded β-sandwich of jelly-roll topology (Shimon *et al.*, 1997). In the present communication, we report the cloning, overexpression, purification, crystallization and preliminary X-ray analysis of cohesin AF2375 from *A. fulgidus*.

2. Cloning, expression and purification

The DNA encoding the cohesin module from AF2375 (NP_071198; residues 273–433; AFcoh2375) was amplified by the polymerase chain

Table 1
Crystal parameters and data-collection statistics.

Values in parentheses are for the highest resolution shell.

X-ray source	ESRF, ID29
Wavelength (Å)	0.934
Temperature (K)	100
Detector	ADSC CCD, Q210
Space group	$P4_332$
Unit-cell parameters (Å)	$a = b = c = 101.75$
Resolution (Å)	42.2–1.82 (1.90–1.82)
Mosaicity (°)	0.418
V_M (Å^3 Da^{-1})	2.48
Solvent content (%)	50.44
Wilson B factor (Å^2)	41.74
No. of measured reflections	268962
No. of unique reflections	16016 (765)
Completeness	99.9 (100.0)
R_{merge}†	0.079 (0.56)
$\langle I/\sigma(I)\rangle$	75.09 (1.20)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\sum_{hkl}$ denotes the sum over all reflections and $\sum_i$ the sum over all equivalent and symmetry-related reflections (Stout & Jensen, 1968).

reaction (PCR) using a single colony of *A. fulgidus* strain ATCC 49558D-5 (DSM 4304). The following PCR oligonucleotide primers were used: forward, 5′-GCCCCATGGGGCTTCCTCCGAAAACTACCATC-3′; reverse, 5′-GCACTCGAGTGCTTCTTCCTGAGAGACAATCCT-3′. The resulting PCR product was cloned into the *Nco*I and *Xho*I sites of the pET28a(+) expression vector (Novagen). The construct included a hexahistidine tag attached to the 5′ end and was expressed in *Escherichia coli* strain BL21 cells grown at 310 K in 4 l Luria–Bertani medium supplemented with 30 µg ml^{-1} kanamycin. Protein expression was induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.5 and cultivation was continued at 289 K overnight. Cells were harvested by centrifugation (4000*g* for 15 min) at 277 K and resuspended in 50 m*M* NaH_2PO_4 pH 8.0 containing 300 m*M* NaCl at a ratio of 1 g wet pellet to 4 ml buffer solution. DNase was added prior to the sonication procedure. The suspended pellet was sonicated in a Sonicator Ultrasonic Processor X2 (Misonix Inc.) for 20 min in discontinuous mode (0.05 s pulse on and 0.05 s pulse off). The suspension was kept on ice during sonication; it was then centrifuged (20 000*g* at 277 K for 20 min) and the supernatant fluids were collected. The expressed His-tagged protein was isolated by metal-chelate affinity chromatography using Ni–IDA resin (Rimon Biotech, Israel). The wash and elution buffers used were 50 m*M* sodium phosphate buffer pH 6.0, 300 m*M* NaCl, 10% glycerol and 50 m*M* sodium phosphate buffer pH 6.0, 300 m*M* NaCl, 10% glycerol, 250 m*M* imidazole, respectively. No attempts were made to remove the His tag. Fast protein liquid chromatography (FPLC) was performed using a Superdex 75 pg column and an ÄKTA Prime system (GE Healthcare, Haifa, Israel) in column buffer (50 m*M* Tris pH 7.5, 0.15 *M* NaCl, 0.05% sodium azide). The protein was concentrated to 10 mg ml^{-1} using Centriprep YM-3 centrifugal filter devices (Amicon Bioseparation, Millipore Inc., Billerica, USA). The protein concentration was determined by measuring the absorbance at 280 nm using the calculated extinction coefficient of the protein (ε_{280} = 0.247 g l^{-1}). The final degree of purity of the protein was around 95%, as estimated from SDS–PAGE (not shown).

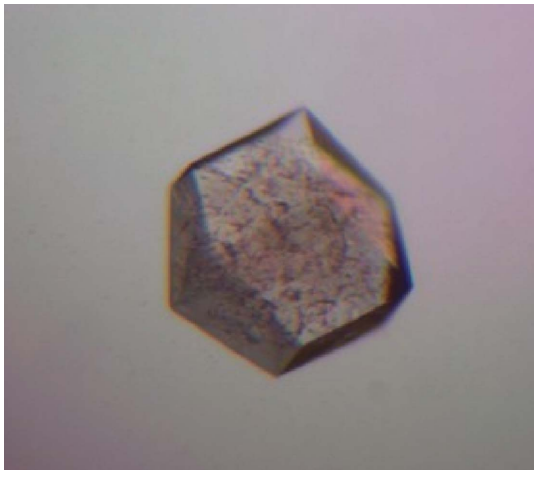


Figure 1
A crystal of AFcoh2375 from *A. fulgidus* grown in 0.1 *M* Tris–HCl pH 8.5, 2 *M* mono-ammonium dihydrogen phosphate.

3. Crystallization

Crystallization was performed using Hampton Research Crystal Screen with the hanging-drop method at 293 K. AFcoh2375 crystals (Fig. 1) were obtained after one month in a 4 µl drop containing 2 µl protein solution and 2 µl reservoir solution (0.1 *M* Tris–HCl pH 8.5, 2 *M* mono-ammonium dihydrogen phosphate) equilibrated against 0.4 ml reservoir solution.

4. X-ray diffraction

Crystals were harvested from the crystallization drop surrounded by mother liquor using thin-walled glass capillaries (Glas Technik & Konstruktion, Berlin). The capillaries were sealed at the narrow end using a flame and at the funnel end with high vacuum grease to facilitate transfer to the synchrotron. Diffraction data were collected on the ID29 beamline at the ESRF (European Synchrotron Radiation Facility, Grenoble, France). X-ray radiation of 0.934 Å wavelength and a CCD ADSC detector were used. The crystals, which were dismounted from the capillary into a solution mimicking the mother liquor with the addition of 25% ethylene glycol, were incu-

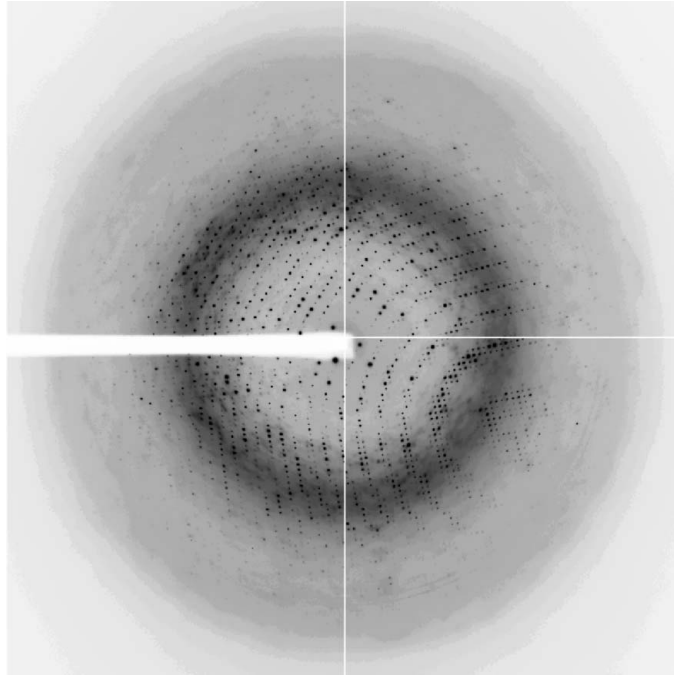


Figure 2
Diffraction pattern from the crystal of AFcoh2375 measured with 0.5° oscillation.

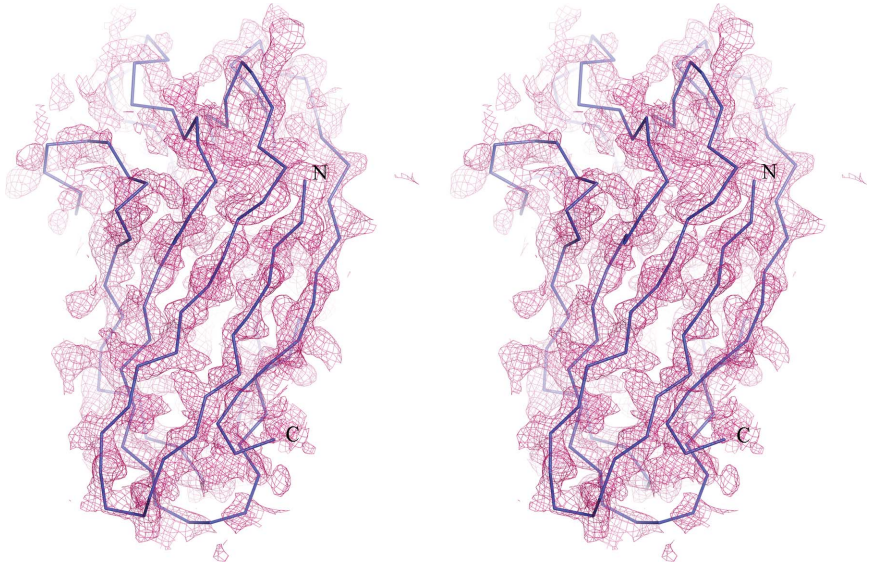


Figure 3
Stereoview rendered using *PyMOL* (DeLano, 2002) of a slice through the superposition of one of the model molecules (PDB code 1anu) used to solve the structure and $2F_o - F_c$ likelihood-weighted electron density after rigid-body refinement.

bated for less than 1 min, mounted in a cryoloop (Teng, 1990) and placed in a stream of cold nitrogen at 100 K generated by an Oxford Cryosystem (Cosier & Glazer, 1986). The data set was collected in 0.5° oscillation frames and was indexed, processed and scaled using *DENZO* and *SCALEPACK* as implemented in *HKL*-2000 (Otwinowski & Minor, 1997). The crystals diffracted to 1.82 Å resolution (Fig. 2) and belonged to the primitive cubic space group *P*432, with unit-cell parameters $a = b = c = 101.75$ Å. Systematic absences suggested two possible space groups: $P4_132$ or $P4_332$. According to the sequence composition of the protein molecule, the Matthews density V_M (Matthews, 1968) was 2.48 Å^3 Da^{-1}, assuming the presence of one molecule in asymmetric unit, with a solvent content of 50.44%. Diffraction data statistics are presented in Table 1.

5. Molecular replacement

A search ensemble of models for molecular replacement was constructed from six pre-aligned structures of cohesin molecules (PDB codes 1anu, 1aoh, 1g1k, 1qzn, 1zv9 and 1tyj). The highest sequence identity between AFcoh2375 and the listed cohesins was 20%. The r.m.s.d.s between the cohesin molecules used for molecular replacement ranged from 1.6 to 2.5 Å. After attempts to solve the structure by molecular-replacement methods based on a search in Patterson space, the structure was solved by a six-dimensional real-space search using the *BEAST* molecular-replacement program (Read, 2001). Large differences in sequence homology between AFcoh2375 and the model structures were reflected in the relatively high R and R_{free} factors at the initial stage of rigid-body refinement: 0.57 and 0.54, respectively. Such values may have indicated that the solution was wrong; however, the electron-density map calculated after rigid-body refinement was convincing and exhibited features that were in partial accordance with the AFcoh2375 sequence (Fig. 3). Automatic building of the structure was attempted with *ARP*/*wARP* (Perrakis *et al.*, 1999) and succeeded in tracing 80 of the 160 residues of AFcoh2375. *ARP*/*wARP* was implemented once again, this time in structure-improvement mode (Lamzin & Wilson, 1993), resulting in a hybrid structure that consisted of 80 amino-acid residues of the protein molecule and a large group of dummy atoms that were modeling electron density unexplained by the present structure. The hybrid model was refined using *REFMAC*5 (Murshudov *et al.*, 1997) and several gaps in the loops of the structure were manually rebuilt using the graphical programs *O* (Jones *et al.*, 1991) and *Coot* (Emsley & Cowtan, 2004). *ARP*/*wARP* was implemented for a third time, again in automatic model-building mode (Perrakis *et al.*, 1999), this time tracing and docking 122 of 160 residues. Cycling of manual and automatic rebuilding and refinement of the structure is presently in process.

We acknowledge the ESRF for synchrotron beam time and thank the staff scientists of the ID14 beamline cluster for their assistance. This research was supported by the Israel Science Foundation (grant Nos. 422/05, 159/07 and 291/08) and by grants from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. EAB holds The Maynard I. and Elaine Wishner Chair of Bio-Organic Chemistry.

References

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr. Opin. Struct. Biol.* **8**, 548–557.
Bayer, E. A., Coutinho, P. M. & Henrissat, B. (1999). *FEBS Lett.* **463**, 277–280.
Bayer, E. A., Morag, E. & Lamed, R. (1994). *Trends Biotechnol.* **12**, 379–386.
Chitayat, S., Gregg, K., Adams, J. J., Ficko-Blean, E., Bayer, E. A., Boraston, A. B. & Smith, S. P. (2008). *J. Mol. Biol.* **375**, 20–28.
Cosier, J. & Glazer, A. M. (1986). *J. Appl. Cryst.* **19**, 105–107.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Haimovitz, R., Barak, Y., Morag, E., Voronov-Goldman, M., Shoham, Y., Lamed, R. & Bayer, E. A. (2008). *Proteomics*, **8**, 968–979.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Salamitou, S., Tokatlidis, K., Beguin, P. & Aubert, J. P. (1992). *FEBS Lett.* **304**, 89–92.
Shimon, L. J. W., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y. & Frolow, F. (1997). *Structure*, **5**, 381–390.
Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide.* London: MacMillan.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Yaron, S., Morag, E., Bayer, E. A., Lamed, R. & Shoham, Y. (1995). *FEBS Lett.* **360**, 121–124.