Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, expression, crystallization and preliminary X-ray analysis of the DNA-binding protein Sso10a from *Sulfolobus solfataricus*

M. J. Teale,[a]* M. Kahsai,[a] S. K. Singh,[b] S. P. Edmondson,[a] R. Gupta,[b] J. W. Shriver[a] and E. Meehan[a]

[a]Laboratory for Structural Biology, Chemistry Department, University of Alabama in Huntsville, Huntsville, AL 35899, USA, and [b]Department of Biochemistry and Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, IL 62901, USA

Correspondence e-mail: tealem@email.uah.edu

Received 21 March 2003
Accepted 13 May 2003

The gene for the DNA-binding protein Sso10a from the hyperthermophilic archaeon *Sulfolobus solfataricus* was cloned and overexpressed in *Escherichia coli*. Crystals of the purified protein have been grown that diffract to beyond 2.15 Å resolution. The protein crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 57.24$, $b = 60.16$, $c = 69.96$ Å. With one dimer per asymmetric unit, the crystal to volume per protein mass (V_M) is 2.9 Å^3 Da^{-1} and the solvent content is ∼57%. Complete X-ray diffraction native data were collected from a single crystal and processed to 2.15 Å.

1. Introduction

The chromatin of the crenarchaeon *Sulfolobus* contains a number of small basic relatively abundant DNA-binding proteins (Grote *et al.*, 1986). These have been classified into three groups according to their molecular weights (Dijk & Reinhardt, 1986). The 7 kDa class is composed of two to five homologous species (*e.g.* Sso7d and Sso7e in *S. solfataricus* and Sac7a, Sac7b, Sac7c, Sac7d and Sac7e in *S. acidocaldarius*). The structure and thermodynamics of Sac7d and Sso7d have been the most thoroughly studied of the *Sulfolobus* chromatin proteins (Choli *et al.*, 1988; Baumann *et al.*, 1994; Knapp *et al.*, 1996; Agback *et al.*, 1998; Gao *et al.*, 1998; Lundbäck *et al.*, 1998; Mai *et al.*, 1998; Robinson *et al.*, 1998; Shriver *et al.*, 2001; Edmondson & Shriver, 2001) and these are the only *Sulfolobus* chromatin proteins for which a protein–DNA complex structure is known (Robinson *et al.*, 1998; Gao *et al.*, 1998). Little is known about the 8 kDa class except that it is apparently composed of two species (Sso8a and Sso8b). The 10 kDa group is composed of three unrelated species (Sso10a, Sso10b and Sso10b2; Dijk & Reinhardt, 1986; Forterre *et al.*, 1999) and only Sso10b has been studied beyond the delineation of primary sequences. The DNA-binding affinity of Sso10b is affected by acetylation and has been renamed Alba (acetylation lowers binding affinity) by Bell *et al.* (2002). An X-ray crystal structure of Alba has been recently published recently along with a proposed model for the DNA–protein complex (Wardleworth *et al.*, 2001, 2002).

Although all these proteins have been loosely described as histone-like proteins, no homology to archaeal or eukaryotic histones exists, nor is there any homology to the bacterial HU proteins. True histones exist in a number of euryarchaeota (Sandman & Reeve, 2000), but there is no evidence to date for such proteins in the crenarchaeota (which include *Sulfolobus*, *Acidianus*, *Desulfurococcus*, *Pyrodictium* and *Pyrobaculum*). Chromatin regulation in the crenarchaeota appears to resemble that in the eukaryotes in that gene transcription is regulated by a modification of chromatin structure (White & Bell, 2002). For example, Sir2, a protein which can deacetylate histones, can also deacetylate Alba (Sso10b) and increase its DNA affinity (Bell *et al.*, 2002). It has been argued that chromatin structure-regulating mechanisms may have emerged before the separation of the archaea and eukaryotes (White & Bell, 2002).

The archaeal chromatin structure and its regulation may provide insights into the mechanisms of chromatin regulation in eukaryotes. Specifically, we are interested in characterizing the structures of the chromatin proteins of *Sulfolobus* and describing their interactions quantitatively. Toward this end, we present here the expression, crystallization and initial X-ray analysis of Sso10a.

2. Materials and methods

2.1. Expression and purification of Sso10a

Native *S. acidocaldarius* Sac10a protein was purified from *S. acidocaldarius* using methods similar to those for Sac7d (Edmondson & Shriver, 2001) and was identified according to the criteria of Grote *et al.* (1986) as a basic 10 kDa protein (based on SDS–gel electrophoresis and MALDI-TOF mass spectrometry) eluting prior to Sac7d on a cation-exchange column. Specifically, acid-soluble proteins from *S. acidocaldarius* were purified on HP-SP Sepharose (Pharmacia) with a linear 0–0.5 *M* NaCl gradient in 0.01 *M* KH_2PO_4

pH 7. The protein eluting prior to Sac7d at 0.2 *M* NaCl was further purified by reverse-phase HPLC using a Resource 15RPC column (Pharmacia) with a linear acetonitrile/TFA gradient. A more detailed description of native Sac10a and its DNA-binding affinity will be presented elsewhere.

No sequence information was reported for Sac10a by Grote *et al.* (1986) owing to a blocked amino-terminus, which was confirmed by this laboratory. Sequence analysis of three tryptic fragments (Fig. 1) of the native protein were used to search the *S. solfataricus* P2 genome database (http://www-archbac.u-psud.fr/projects/sulfolobus/) for the homologous *Sso10a* gene (annotated as SSO10449 in Fig. 2). The *Sso10a* gene was amplified by PCR from *S. solfataricus* genomic DNA [kindly supplied by Dr Dennis Grogen (University of Cincinnati) from a *S. solfataricus* strain colony purified from the P2 strain in the laboratory of Dr W. Zillig (Regensburg)]. The sequence of the amplified gene revealed several differences between the cloned gene and the P2 databank sequence. All but one of the differences were substitutions of lysine for arginine or arginine for lysine. The amplified gene also has an additional glutamine at the C-terminus (Fig. 1). A *BLAST* search showed homology to several ORFs in other archaea that belong exclusively to the euryarchaeota (Fig. 2). No sequence homology was noted to Sso10b or Ssh10b2.

The gene was cloned into the expression vector pETBlue-2 (Novagen). Transformation was effected into RossettaBlue (DE3)pLacI cells (Novagen). Cells were grown at 310 K and protein expression was carried out at 302 K in standard LB medium. Cultures induced with 1 m*M* ITPG on reaching an optical density of 0.9 at 600 nm were harvested after 8 h by centrifugation and stored at 193 K until required.

The frozen cells were thawed and resuspended in ice-cold Tris–EDTA buffer (10 m*M* Tris–HCl pH 8.00, 10 m*M* EDTA pH 8.00, 0.1% Triton X-100, 0.5 m*M* PMSF) and lysed by sonication. DNAse I treatment was carried out at 315 K. The lysate was then incubated in a 343 K water bath for 30–40 min to denature and precipitate *Escherichia coli* proteins. The crude extract was centrifuged at a maximal relative centrifugal force of 300 000*g* in a Sorvall Discovery 90SE using a T1250 rotor and saving the supernatant. Filtered supernatant (0.45 µm filter) was then loaded onto an Amersham Hi-trap SP cation-exchange column previously equilibrated with 10 m*M* KH_2PO_4 pH 7.0 (buffer *A*) and eluted with a linear gradient of 0–1.0 *M* NaCl at room temperature. Protein peaks were analyzed using MALDI–TOF mass spectrometry and SDS–PAGE techniques, which confirmed the expected molecular weight of the protein (data not shown). The identity of the Sso10a gene product was confirmed by 20 cycles of N-terminal sequencing of the expressed protein. A typical yield of protein per litre of culture is 13 mg.

```
      KRTRYEI  IHDIL          KTWLMYK
                                 K ANLSYELTNN Y
1  MAKKRSKLEI IQAILEACKS GSPKTRIMYG ANLSYALTGR YIKMLMDLEI 50
   MAKKKSKLEI IQAILEACKS GSPKTRIMYG ANLSYALTGR YIKMLMDLEI

51 IKQEGKQYML TKKGEELLED IRRFNDMRKN MDQLKEKINS VLSIK      95
   IRQEGKQYML TKKGEELLED IRKFNEMRKN MDQLKEKINS VLSIRQ
```

Figure 1
Manual alignment of the three tryptic fragment sequences (underlined) to the *S. solfataricus* P2 database sequence (upper sequence) and the PCR-amplified gene (lower sequence). Differences in the two protein sequences derived from the nucleotide sequences are shown in bold.

```
AF2083          --------------MGDKRSRFEIFVAILKITN---SGANMTKIVYGANINFKMAQSYLE
AF1663          -------------MAEMRRSKIEIMAEILQTCSG--NGVNITKIVYRVNLNSKVAQDYIN
SSO10449        --------------MAKKKSKLEIIQAILEACK---SGSPKTRIMYGANLSYALTGRYIK
MJ0287          MRIPPPLSNNKNRNRRYKRSQFEVIFEILHIIKE--GEQIKTRIMYAANLDWRNFSKYID
Ta0385          -----MESLMGIGPTRKNRDRGDLVFSILKELYISDGQIGITRLIYKVNTNYVIAKDILE
                              .:.: ::.  **.      .    *:::* .* .       :.

AF2083          ELVENDFLVLDSKNGKKTYSTTDKGRDFVKKYGELQEMTE--------------
AF1663          CLLKEGFIEPVTDGKRTKYRTTEKGREFIQKFSEIEEDLQALKNAVENSIF---
SSO10449        MLMDLEIIRQEGK-QYMLTKKGEELLEDIRKFNEMRKNMDQLKEKINSVLSIRQ
MJ0287          FLISNGFIKKNKE-KFELTELGKKLYSSLYELFEIMNSKP--------------
Ta0385          MLSGSGLIENQAKNERNGFKVTEKGLRFIEAYERINSILGFKRI----------
                 *    ::    .         .:    :    .: .
```

Figure 2
Alignment of theoretical proteins homologous to Sso10a (SS010449) using the *BLAST*2 program (Altschul *et al.*, 1997). Genes with an AF prefix belong to the organism *Archaeoglobus fulgidus*, SS = *S. solfataricus*, MJ = *Methanocaldococcus jannaschii*, Ta = *Thermoplasma acidophilum*.

Table 1
Processing statistics for X-ray diffraction data.

Values in parentheses refer to the highest resolution bin.

Resolution range (Å)	15–2.15 (2.23–2.15)
Total observations	365374
Unique reflections	13841
Completeness (%)	96.0 (100)
R_{merge}† (%)	13 (64)
$I/\sigma(I)$	10 (3)
Space group	$P2_12_12_1$
Matthews coefficient (Å^3 Da^{-1})	2.9
Solvent content (%)	57

† $R_{merge} = \sum I - \langle I \rangle / \sum \langle I \rangle$.

2.2. Crystallization and X-ray data collection

Crystallization trials were carried out using the hanging-drop vapor-diffusion method of crystallization, Linbo boxes and various Hampton Crystal Screens (Hampton Research, California) at 295 K. X-ray diffraction data were collected using a Rigaku rotating copper-anode X-ray generator at 48 kV and 98 mA (λ = 1.5418 Å) with an R-AXIS IV image plate. When collecting X-ray diffraction data, crystals were taken directly from the drop using a 0.1 mm Hampton cryoloop, submerged in liquid nitrogen and then transferred to a nitrogen cold stream at approximately 93 K. A total of 150 rotation images were collected at a detector distance *d* of 180 mm and a rotation angle φ of 1°.

3. Results

Initially, several conditions were found to produce crystals of various habits using the Hampton Crystal Screens. Many of these crystals diffracted strongly, but they all demonstrated a pseudo-fiber diffraction pattern which we have not indexed at this time. Hampton PEG/Ion screen condition Nos. 1, 2 and 3 produced similar crystals that diffracted in the pseudo-fiber diffraction pattern, but continued refinement of condition No. 1 (0.2 *M* Li_2SO_4, 20% PEG 3350) produced crystals suitable for diffraction data collection. Even so, most crystals from these conditions exhibited a pseudo-fiber diffraction pattern when X-ray data were collected. The final crystallization conditions were 95–85% of a saturated Li_2SO_4 solution in 24% PEG 3350 at 295 K. The crystals nucleate overnight, grow slowly for 10–14 d and reach dimensions of approximately 0.15 × 0.15 × 0.05 mm; they diffract to beyond 2.15 Å. The crystal which produced

the usable diffraction data appeared to nucleate on the cover slip. Complete native X-ray diffraction data were collected from a single crystal and processed to 2.15 Å (Table 1).

4. Discussion

The protein Sso10a is a DNA-binding protein implicated to have histone-like properties, yet little is known about its structure and function. The determination of this structure may improve our understanding of DNA stability and utility in the hyperthermophilic archaea. The crystallization of Sso10a in a suitable crystal habit has proven to be arduous owing to its tendency to produce crystals that yield a pseudo-fiber-like diffraction pattern. Obtaining phases may prove to be equally difficult especially since we know of no homologous structures that might be used to obtain molecular-replacement phases. Direct methods for obtaining phase information may not yield a clean solution given that there are eight S atoms per protein and probably two monomers per asymmetric unit, which yields 16 individual sulfur positions in the asymmetric unit. However, this technique may yet prove to be successful if combined with selenomethionine-replacement experiments for two reasons. Firstly, it is less complex to find 14 Se atoms than 16 S atoms; also, Se atoms give a stronger anomalous signal than the S atoms they would replace. In a recent visit to the SER-CAT ID beamline at the Argonne National Laboratory wavelength scans were run on possible derivatives developed using the technique of co-crystallization. It was found that we have probably co-crystallized Sso10a with a platinum derivative, but these crystals were too small to collect useful data (data not shown). We are currently attempting to grow larger crystals of the same type and continue to screen for other heavy-atom derivatives.

This work was in part supported by NIH grants GM49686 (JWS) and GM55945 (RG) and by a generous gift from an anonymous donor to the Laboratory for Structural Biology, University of Alabama in Huntsville. The authors are thankful to Joyce Looger for her support with computers and computer programs.

References

Agback, P., Baumann, H., Knapp, S., Ladenstein, R. & Hard, T. (1998). *Nature Struct. Biol.* **7**, 579–584.

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

Baumann, H., Knapp, S., Lundback, T., Ladenstein, R. & Hard, T. (1994). *Nature Struct. Biol.* **11**, 808–819.

Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P. & White, M. F. (2002). *Science*, **296**, 148–151.

Choli, T., Henning, P., Wittmann-Liebold, B. & Reinhardt, R. (1988). *Biochim. Biophys. Acta*, **950**, 193–203.

Dijk, J. & Reinhardt, R. (1986). *Bacterial Chromatin*, edited by C. O. Gualerzi & C. L. Pon, pp. 185–218. Berlin: Springer–Verlag.

Edmondson, S. P. & Shriver, J. W. (2001). *Methods Enzymol.* **334**, 129–145.

Forterre, P., Confalonieri, F. & Knapp, S. (1999). *Mol. Microbiol.* **32**, 669–670.

Gao, Y. G., Su, S. Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B. S., Edmondson, S. P., Shriver, J. W. & Wang, A. H. (1998). *Nature Struct. Biol.* **9**, 782–786.

Grote, M., Dijk, J. & Reinhardt, R. (1986). *Biochim. Biophys. Acta*, **873**, 405–413.

Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B. & Ladenstein, R. (1996). *J. Mol. Biol.* **264**, 1132–1144.

Lundbäck, T., Hansson, H., Knapp, S., Ladenstein, R. & Hard, T. (1998). *J. Mol. Biol.* **276**, 775–786.

Mai, V. Q., Chen, X., Hong, R. & Huang, L. (1998). *J. Bacteriol.* **180**, 2560–2563.

Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. & Wang, A. H. (1998). *Nature (London)*, **392**, 202–205.

Sandman, K. & Reeve, J. N. (2000). *Arch. Microbiol.* **173**, 165–169.

Shriver, J. W., Peters, W. B., Szary, N., Clark, A. T. & Edmondson, S. P. (2001). *Methods Enzymol.* **334**, 389–422.

Wardleworth, B. N., Russell, R. J. M., Bell, S. D., Taylor, G. L. & White, M. F. (2002). *EMBO J.* **21**, 4654–4662.

Wardleworth, B. N., Russell, R. J. M., White, M. F. & Taylor, G. L. (2001). *Acta Cryst.* D**57**, 1893–1894.

White, M. F. & Bell, S. D. (2002). *Trends Genet.* **18**, 621–626.