Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

Expression, purification, crystallization and preliminary X-ray crystallographic analysis of the histone-like HU protein from *Spiroplasma melliferum* KC3

Konstantin Boyko,[a,b]* Marina Gorbacheva,[a,b] Tatiana Rakitina,[b,c] Dmitry Korzhenevskiy,[b] Anna Vanyushkina,[d] Dmitry Kamashev,[d] Alexey Lipkin[b] and Vladimir Popov[a,b]

[a]Laboratory of Enzyme Engineering, A. N. Bach Institute of Biochemistry, RAS, Leninsky Prospekt 33/2, Moscow 119071, Russian Federation, [b]The Protein Factory, NBICS Center, National Research Centre 'Kurchatov Institute', Akad. Kurchatova Square 1, Moscow 123182, Russian Federation, [c]Laboratory of Hormonal Regulation Proteins, Institute of Bioorganic Chemistry, RAS, Miklukho-Maklaya 16/10, Moscow 117997, Russian Federation, and [d]SRI of Physical-Chemical Medicine, Malaya Pirogovskaya 1a, Moscow 119435, Russian Federation

Correspondence e-mail: kmb@inbi.ras.ru

Received 7 October 2014
Accepted 18 November 2014

© 2015 International Union of Crystallography
All rights reserved

HU proteins belong to the nucleoid-associated proteins (NAPs) that are involved in vital processes such as DNA compaction and reparation, gene transcription *etc.* No data are available on the structures of HU proteins from mycoplasmas. To this end, the HU protein from the parasitic mycoplasma *Spiroplasma melliferum* KC3 was cloned, overexpressed in *Escherichia coli* and purified to homogeneity. Prismatic crystals of the protein were obtained by the vapour-diffusion technique at 4°C. The crystals diffracted to 1.36 Å resolution (the best resolution ever obtained for a HU protein). The diffraction data were indexed in space group *C*2 and the structure of the protein was solved by the molecular-replacement method with one monomer per asymmetric unit.

1. Introduction

HU proteins are the most abundant DNA-binding proteins in prokaryotic organisms, where they play a substantial role in the processes of DNA repair, recombination and replication (Drlica & Rouviere-Yaniv, 1987; Kamashev *et al.*, 2008). HU belongs to the nucleoid-associated proteins (NAPs) involved in DNA compaction in the bacterial nucleoid and in the regulation of the main DNA transactions, including gene transcription. In many bacteria HU is the only protein with a NAP function, and its deletion from the bacterial genome is lethal in many cases (Li & Waters, 1998; Sasaki *et al.*, 2009). These small (about 90 amino acids per monomer) basic proteins are annotated in 99% of bacteria with sequenced genomes, and several thousand HU sequences are known (Drlica & Rouviere-Yaniv, 1987). The three-dimensional structures of a number of HU proteins, as well as of their mutants, have been solved. These proteins form homodimers or heterodimers fastened by hydrogen bonds and salt bridges as well as large central hydrophobic interactions (Christodoulou *et al.*, 2003; Kawamura *et al.*, 1998; Ramstein *et al.*, 2003; White *et al.*, 1999). Moreover, many of the proteins possess high thermotolerance (Kawamura *et al.*, 1998; Orfaniotou *et al.*, 2009; Ramstein *et al.*, 2003). Thus, HU proteins are convenient models for investigation of the structural basis of thermal stability.

Mycoplasmas are the smallest known microorganisms, with a parasitic lifestyle and drastically reduced genome sizes (Gorbachev *et al.*, 2013). Structural data on HU proteins from mycoplasmas are lacking. The HU protein from *Spiroplasma melliferum* KC3 (HUSpm), an insect parasite infecting honey bees (Schwarz *et al.*, 2014), was cloned, overexpressed in *Escherichia coli* and purified. Despite the fact that HUSpm originates from a mesophilic organism, it possesses a unique thermal stability (K. Boyko, to be published) comparable to those of HU proteins from thermophiles. The protein was successfully crystallized and its structure was solved at 1.36 Å resolution.

2. Materials and methods

2.1. Macromolecule production

The single gene encoding the histone-like DNA-binding protein HU (*hup2*; start 30510, end 30226) annotated in the *S. melliferum* KC3 genome (Alexeev *et al.*, 2012) was amplified from genomic DNA by PCR with primers generated for directed cloning into the multiple cloning site (Table 1). Following *Nde*I and *Eco*RI digestion, the PCR

Table 1
Macromolecule-production information.

Source organism	*S. melliferum* KC3
DNA source	*S. melliferum* KC3
Forward primer†	GGTGTACATATGTCAAAAAAAGAACTAGC
Reverse primer†	CTTTCGGAATTCTTAATTATTGTTTAAATCAG
Cloning vector	pET-21d
Expression vector	pET-21d modified by an N-terminal His_6 tag followed by a TEV protease digestion site
Expression host	*E. coli* BL21-CodonPlus (DE3)-RIPL
Complete amino-acid sequence of the construct produced‡	MGSDKI*HHHHHH*ENLYFQGHMSKKELAAQIAEKFTDVLSKTHFQGHMSKKELAAQIAEKFTDVLSKTHAEEITNFVFDHIKKALVAGKEVSIAGFGKFAVTERAARDGRNPSTGETIKIPASKSAKFKAGKQLKTDLNNN

† Restriction sites are underlined. ‡ The N-terminal His_6 tag is shown in italics and the TEV protease digestion site is underlined.

product was ligated in frame with the oligonucleotide duplex encoding the N-terminal tag sequence (Table 1) into the pET-21d cloning vector (5443 bp; Novagen, Darmstadt, Germany) digested with the *Nde*I and *Eco*RI restriction enzymes. The duplex was produced by annealing the oligonucleotides containing 5′ overhangs complementing the *Nco*I and *Nde*I restriction sites: *Nco*I_His6TEV_for (5′-CATGGGATCTGATAAAATTCATCATCATCATCACCACGAAAACCTGTACTTCCAGGGCCA-3′) and *Nde*I_His6TEV_rev (5′-TATGGCCCTGGAAGTACAGGTTTTGTGGTGATGATGATGATGAATTTTATCAGATCC-3′). The construct was verified by sequencing and transformed by heat shock into *E. coli* BL21-CodonPlus (DE3)-RIPL competent host cells (Stratagene, La Jolla, USA) containing the chloramphenicol-resistant pRARE2 plasmid, which supplied seven rare tRNAs. The *E. coli* preculture was grown overnight on Luria–Bertani (LB) medium containing 100 mg ml^{-1} ampicillin and 34 mg ml^{-1} chloramphenicol at 310 K and used as a 1%(*v*/*v*) inoculum for 4 l LB–ampicillin expression cultures. The cultures were incubated at 310 K until an OD_{600} of 0.8 was reached and were then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 1 m*M*. After incubation for an additional 18 h at 297 K, the cells were harvested by centrifugation. The cell pellet was stored at 203 K until further use.

For the Ni–NTA affinity purification of His_6TEV-HUSpm, the cells were thawed on ice and resuspended in lysis buffer (50 m*M* Tris–HCl, 500 m*M* NaCl, 10 m*M* imidazole, 5% glycerol, 0.2% Triton X-100 pH 8.0 supplemented with 1 m*M* PMSF). The cells were disrupted by sonication on ice (40%, 5 min; UP200S ultrasonication processor, 24 kHz, 200 W; Dr Hielscher GmbH, Teltow, Germany). Cell debris was removed from the crude extract by centrifugation at 25 000*g* for 30 min at 277 K. The cleared extract was loaded onto a 5 ml Ni–NTA Superflow column (Qiagen, Netherlands) equilibrated with the same buffer. The nonspecifically bound proteins were removed by washing with five column volumes (CV) of lysis buffer and 5 CV wash buffer (50 m*M* Tris–HCl, 1 *M* NaCl, 40 m*M* imidazole pH 8.0). The His_6-tagged protein was eluted with 3 CV elution buffer [50 m*M* Tris–HCl, 500 m*M* NaCl, 300 m*M* imidazole, 5%(*v*/*v*) glycerol] and protein-containing fractions were selected based on their UV absorbance at 214 nm. The Ni–NTA purified fractions were pooled and incubated overnight at 277 K with TEV protease at a molar ratio of about 10:1 and then analyzed on 15% SDS–PAGE stained with Coomassie Brilliant Blue. Following the removal of imidazole by dialysis, the pooled fractions were passed again through the 5 ml Ni–NTA Superflow column and the flowthrough containing His_6-free HUSpm was collected. The flowthrough was concentrated in centrifugal concentrators (Amicon Ultra 3 kDa cutoff, Millipore Ltd, Ireland) and subjected to final polishing and buffer exchange by size-exclusion chromatography (GE Superdex G75 10/300 GL column on an ÄKTAexplorer liquid-chromatography system; GE Life Sciences, USA) in final buffer [40 m*M* Tris–HCl, 200 m*M* NaCl, 5%(*v*/*v*) glycerol pH 8.0; Fig. 1*a*]. The purity of the HUSpm-containing fractions was estimated by SDS–PAGE with Coomassie staining (Fig. 1*b*) and the protein concentration was measured by the Bradford assay (Sigma Bradford reagent, Sigma–Aldrich, USA). The purest fractions were pooled and concentrated by centrifugation to a final protein concentration of 14 mg ml^{-1}. The macromolecule-production information is listed in Table 1.

2.2. Crystallization

Initial crystallization experiments were carried out with six commercial crystallization screens from Hampton Research (Index HT, Crystal Screen HT, Crystal Screen Cryo HT, PEG/Ion HT, PEGRx HT and SaltRx HT) at the Protein Factory division of the NBICS Center of the National Research Centre 'Kurchatov Institute'. All initial screens were performed using a protein concentration

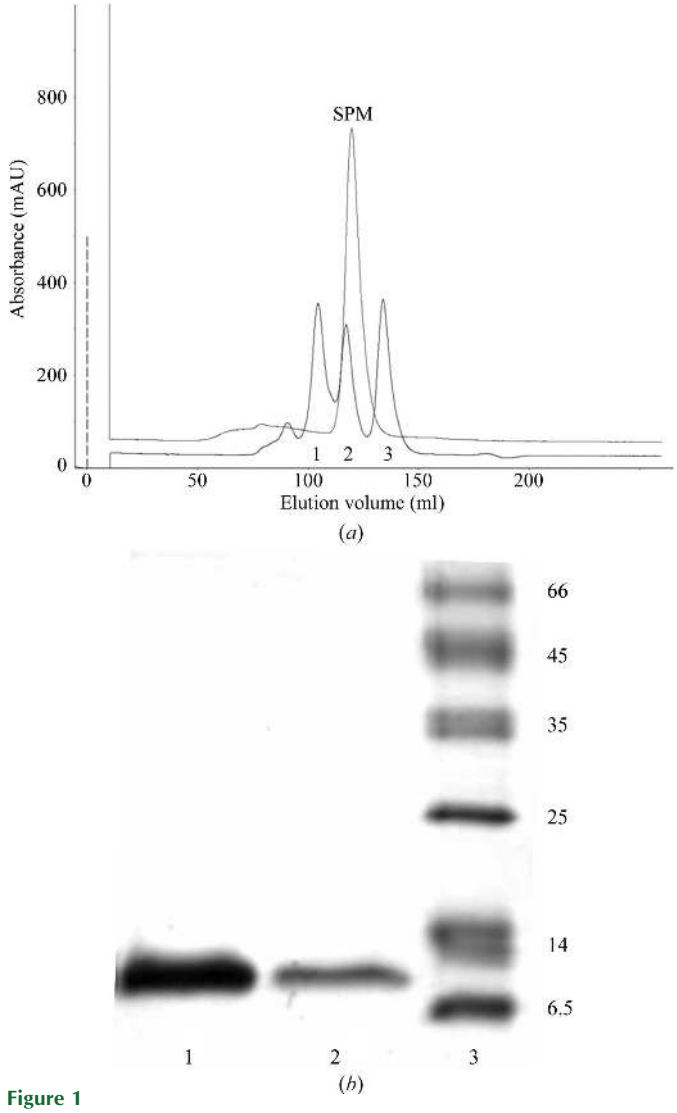

Figure 1
Purification of HUSpm. (*a*) Chromatographic mobility of HUSpm (SPM) on a Superdex G75 10/300 GL column compared with those of three standard proteins (1, ovalbumin, 43 kDa; 2, carbonic anhydrase, 29 kDa; 3, ribonuclease A, 14 kDa). (*b*) Coomassie-stained 15% SDS–PAGE showing the purity of the HUSpm used for crystallization. Lane 1, 0.5 µg HuSpm; lane 2, 0.1 µg HuSpm; lane 3, molecular-weight markers (labelled in kDa).

Table 2
Crystallization.

Method	Hanging-drop vapour diffusion
Plate type	Linbro (24-well)
Temperature (K)	277
Protein concentration (mg ml^{-1})	14
Buffer composition of protein solution	40 m*M* Tris pH 8.0, 200 m*M* NaCl, 5% glycerol
Composition of reservoir solution	0.1 *M* Tris pH 8.0, 35%(*v*/*v*) PEG 400, 5% glycerol
Volume and ratio of drop (protein:precipitant)	300 nl, 1:2
Volume of reservoir (µl)	400

of 14 mg ml^{-1} and the sitting-drop vapour-diffusion method in 96-well CrystalMation plates (Rigaku Automation, USA) at two different temperatures (277 and 293 K). 100 nl protein solution was mixed with an equal volume of reservoir solution and equilibrated against 50 µl reservoir solution. The initial screen resulted in two hits at 277 K in the following conditions: (i) 0.1 *M* HEPES pH 7.0, 30% Jeffamine and (ii) 0.1 *M* Tris–HCl pH 8.0, 30% PEG 400. In condition (i) cracked needle-like crystals appeared in 19 d, but this condition was not reproducible. In condition (ii) plate-like crystals appeared in two months. Owing to their small size, they were unsuitable for a direct X-ray experiment using our equipment. Further optimization of this condition was performed using the hanging-drop vapour-diffusion technique in 24-well VDX plates (Hampton Research, USA) at 277 K by varying the PEG concentration and by the addition of glycerol. Crystals suitable for X-ray diffraction appeared in 1.5 months (Fig. 2). The final crystallization information is listed in Table 2.

2.3. Data collection and processing

Crystals of HUSpm were briefly soaked in mother liquor containing 15% glycerol as a cryoprotectant. Crystals were then flash-cooled to 100 K in liquid nitrogen. X-ray diffraction data were collected on the 'Belok' beamline of the National Research Center 'Kurchatov Institute' (Moscow, Russian Federation) using a MAR CCD detector. A total of 201 images were collected using an oscillation range of 1.0°. The data were indexed, integrated and scaled using *XDS* (Kabsch, 2010). The data-collection and processing statistics are summarized in Table 3.

Figure 2
Crystal of the HU protein from the mycoplasma *S. melliferum*. The prism-shaped crystal has dimensions of about 200 × 300 × 50 µm.

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	'Belok', NRC 'Kurchatov Institute'
Wavelength (Å)	0.984
Temperature (K)	100
Detector	MAR CCD
Crystal-to-detector distance (mm)	90
Rotation range per image (°)	1
Total rotation range (°)	201
Exposure time per image (s)	180
Space group	*C*2
a, *b*, *c* (Å)	57.76, 39.53, 39.28
α, β, γ (°)	90, 108.33, 90
Mosaicity (°)	0.412
Resolution range (Å)	30–1.36 (1.50–1.36)
Total No. of reflections	96243 (15176)
No. of unique reflections	17603 (4105)
Completeness (%)	96.7 (89.5)
Average multiplicity	5.5 (3.7)
$\langle I/\sigma(I)\rangle$	20.36 (3.63)
$R_{\rm merge}$† (%)	8.2 (49.6)
$R_{\rm r.i.m.}$† (%)	12.7 (49.9)
Overall *B* factor from Wilson plot (Å^2)	17.0

† Diederichs & Karplus (1997).

3. Results and discussion

The single gene from the *S. melliferum* KC3 genome encoding the histone-like DNA-binding protein HU was cloned and the HUSpm protein was overexpressed in *E. coli* with subsequent purification to homogeneity.

A sequence alignment of HUSpm against the NCBI database confirmed that the protein belongs to the family of HU/IHF proteins. A subsequent multiple sequence alignment of HUSpm against members of the HU protein family with known three-dimensional structures using *ClustalW*2 (Larkin *et al.*, 2007) revealed that HUSpm shares moderate overall sequence identity (45% or lower) and possesses conserved regions, including a significant central hydrophobic core.

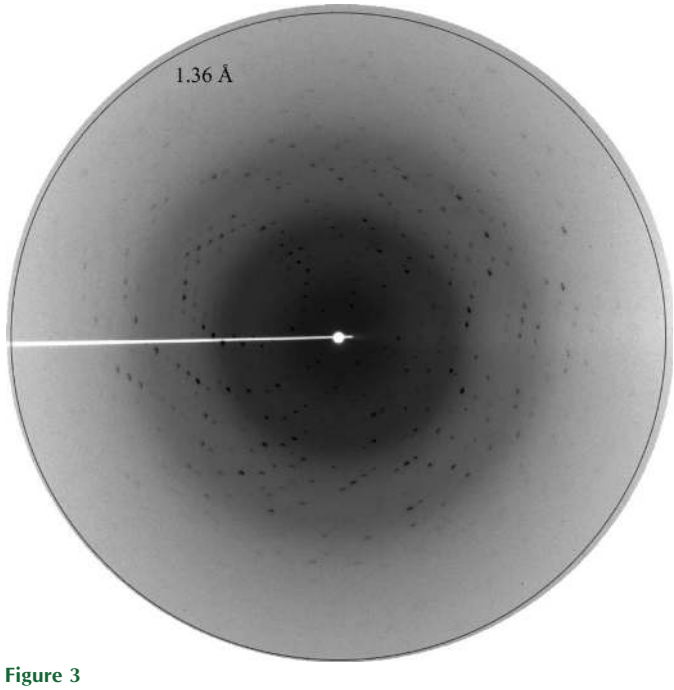

Figure 3
X-ray diffraction from an HUSpm crystal. The diffraction pattern shown was recorded on a MAR CCD detector with 0.984 Å wavelength X-rays, 1° oscillation, 180 s exposure and a sample-to-detector distance of 90 mm.

The *hup2* gene was cloned into the pET-21d plasmid modified by introduction of an oligonucleotide fragment designed for translation of an N-terminal His_6 tag followed by a *Tobacco etch virus* (TEV) protease-digestion site, which allows the purification of the fusion protein based on two subsequent Ni–NTA chromatography steps with TEV protease digestion between them. All purification steps were conducted in a Tris–HCl-based buffer solution. The final size-exclusion chromatography demonstrated the apparent mobility of HUSpm as a 20–25 kDa protein consistent with its homodimerization. According to the SDS–PAGE analysis, we obtained a pure protein with an apparent molecular mass of 11 kDa consistent with its 96-amino-acid size. The yield of purified HUSpm was 8 mg from 1 l culture.

HUSpm was crystallized at a concentration of 14 mg ml^{-1} in a buffer consisting of 40 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl, 5%(*v*/*v*) glycerol. The initial crystallization experiments were performed using commercial screens at the high-throughput crystallization facility (Rigaku, Japan) of the NBICS Center of the NRC 'Kurchatov Institute' ('Protein Factory' division) and resulted in two conditions in which crystals were obtained. The subsequent crystallization-condition optimization gave crystals that diffracted to 1.36 Å resolution (Fig. 3). A data set of 201 images was collected on the 'Belok' synchrotron beamline of the NRC 'Kurchatov Institute'. The crystals of HUSpm belonged to space group *C*2. The Matthews coefficient (Matthews, 1968) strongly indicates that there is one protein monomer per asymmetric unit, with a solvent content of 40%. The structure was solved by the molecular-replacement method using *BALBES* (Long *et al.*, 2008) and the structure of the HU protein from *B. stearothermophilus* (PDB entry 1huu), which shares 44% sequence identity with HUSpm, as the starting model (White *et al.*, 1999). The best solution gave a *Q*-factor of 0.78 with a final *R* factor of 34.7% and R_{free} of 36.1%.

The HUSpm structure has been deposited in the PDB (PDB entry 4n1v). Structural analysis is now in progress, and we expect that it will shed light on the structural basis of the unusual thermal stability of this protein.

This work was supported by the Russian Scientific Fund N14-24-00172 (purification and crystallization as well as structure determination), the Molecular and Cell Biology Program of the Russian Academy of Sciences N6P (cloning and expression) and a Scholarship of the President of Russian Federation for Young Scientists CP-1390.2012.4 (to KMB).

References

Alexeev, D. *et al.* (2012). *J. Proteome Res.* **11**, 224–236.
Christodoulou, E., Rypniewski, W. R. & Vorgias, C. E. (2003). *Extremophiles*, **7**, 111–122.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Mol. Biol.* **4**, 269–275.
Drlica, K. & Rouviere-Yaniv, J. (1987). *Microbiol. Rev.* **51**, 301–319.
Gorbachev, A. Y., Fisunov, G. Y., Izraelson, M., Evsyutina, D. V., Mazin, P. V., Alexeev, D. G., Pobeguts, O. V., Gorshkova, T. N., Kovalchuk, S. I., Kamashev, D. E. & Govorun, V. M. (2013). *BMC Genomics*, **14**, 726.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
Kamashev, D., Balandina, A., Mazur, A. K., Arimondo, P. B. & Rouviere-Yaniv, J. (2008). *Nucleic Acids Res.* **36**, 1026–1036.
Kawamura, S., Abe, Y., Ueda, T., Masumoto, K., Imoto, T., Yamasaki, N. & Kimura, M. (1998). *J. Biol. Chem.* **273**, 19982–19987.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Li, S. & Waters, R. (1998). *J. Bacteriol.* **180**, 3750–3756.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Orfaniotou, F., Tzamalis, P., Thanassoulas, A., Stefanidi, E., Zees, A., Boutou, E., Vlassi, M., Nounesis, G. & Vorgias, C. E. (2009). *Extremophiles*, **13**, 1–10.
Ramstein, J., Hervouet, N., Coste, F., Zelwer, C., Oberto, J. & Castaing, B. (2003). *J. Mol. Biol.* **331**, 101–121.
Sasaki, N., Hirai, M., Maeda, K., Yui, R., Itoh, K., Namiki, S., Morita, T., Hata, M., Murakami-Murofushi, K., Matsuoka, H., Kita, K. & Sato, S. (2009). *FEBS Lett.* **583**, 1446–1450.
Schwarz, R. S., Teixeira, É. W., Tauber, J. P., Birke, J. M., Martins, M. F., Fonseca, I. & Evans, J. D. (2014). *MicrobiologyOpen*, **3**, 341–355.
White, S. W., Wilson, K. S., Appelt, K. & Tanaka, I. (1999). *Acta Cryst.* D**55**, 801–809.