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

Expression, purification, crystallization and preliminary crystallographic study of the SRA domain of the human UHRF1 protein

Bénédicte Delagoutte,[a] Nada Lallous,[a] Catherine Birck,[a] Pierre Oudet[a] and Jean-Pierre Samama[a,b]*

[a]Département de Biologie et de Génomique Structurales, IGBMC, CNRS/INSERM Université Louis Pasteur, Parc d'Innovation BP 10142, 67404 Illkirch CEDEX, France, and [b]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 Gif-sur-Yvette CEDEX, France

Correspondence e-mail: jean-pierre.samama@synchrotron-soleil.fr

Received 28 March 2008
Accepted 26 August 2008

Human UHRF1 belongs to the unique mammalian family of proteins which contain a SET- and RING finger-associated (SRA) domain. This 180-residue domain has been reported to play key roles in the functions of the protein. It allows UHRF1 to bind methylated DNA, histone deacetylase 1 and DNA methyltransferase 1, suggesting a bridge between DNA methylation and the histone code. No structural data is available for any SRA domain. Native and SeMet-labelled SRA domains of human UHRF1 were overexpressed in *Escherichia coli* cells, purified to homogeneity and crystallized using the hanging-drop vapour-diffusion method. A complete MAD data set was collected to 2.2 Å resolution at 100 K. Crystals of the SeMet-labelled protein belonged to the trigonal space group $P3_221$, with unit-cell parameters $a = b = 53.78$, $c = 162.05$ Å.

1. Introduction

In all eukaryotes, DNA is compacted in the nucleus as chromatin, the building block of which is the nucleosome core particle. It consists of 146 base pairs wrapped around an octamer of core histones formed by two copies each of histones H2A, H2B, H3 and H4. The positively charged amino-terminal tail of each histone protrudes from the nucleosome core particle and is subjected to post-translational modifications that modulate the compactness of the chromatin. Jenuwein and coworkers proposed that specific combinations of modifications of histone-tail residues (namely the 'histone code') modulate gene expression (Jenuwein & Allis, 2001). These covalent modifications act as recognition signals that direct the binding of nonhistone proteins and determine the active or repressed transcriptional state of chromatin (de la Cruz *et al.*, 2005).

The human and mouse UHRF1 proteins (also called ICBP90 and NP95) are the best characterized members of the UHRF family. They contain, in the following order, a ubiquitin-like domain, a PHD finger, an SRA domain and a RING finger, which confers E3 ubiquitin ligase activity. In addition to their critical role in cell proliferation, these nuclear proteins appear to participate in epigenetic events. They play a role in maintaining DNA methylation through complex formation with DNA methyltransferase 1 (Achour *et al.*, 2008; Bostick *et al.*, 2007; Sharif *et al.*, 2007). Mouse UHRF1 has been shown to be a new partner for pericentric heterochromatin replication in a process that involves the regulation of histone H4 deacylation (Papait *et al.*, 2007, 2008).

The SRA domain is mainly found in plant and mammalian proteins. In *Arabidopsis thaliana* it is found in histone methyltransferases, where it is involved in directing DNA methylation to target sequences (Naumann *et al.*, 2005). In mammals it has only been identified to date in the UHRF family, in which it displays several functions. Initially, it was characterized as a new methyl-DNA binding domain (Woo *et al.*, 2007; Johnson *et al.*, 2007), with a preference for binding hemimethylated CpG sites in the case of UHRF1 (Bostick *et al.*, 2007). It has also been implicated in the binding of UHRF1 to histone deacetylase 1 (Unoki *et al.*, 2004) and it has recently been shown that the SRA domain promotes the binding affinity of human UHRF1 to K9-methylated histone H3 (Karagianni *et al.*, 2008).

No structural data have been published to date for any SRA domain and here we present the first report of the expression, purification, crystallization and preliminary X-ray analysis of the SRA domain of the human UHRF1 protein (hUHRF1).

2. Materials and methods

2.1. Cloning, expression and purification

The coding sequence for the SRA domain of hUHRF1 (SRA; residues 408–643) was cloned using Gateway technology into the pHGWA expression vector (Busso *et al.*, 2005), which provides an N-terminal hexahistidine tag plus a TEV cleavage site. After TEV digestion, three residues (GHM) remained at the N-terminus of the SRA protein.

Escherichia coli BL21-pLysS(DE3) cells harbouring the SRA expression plasmid were cultured in LB medium containing 100 µg ml^{-1} ampicillin and 35 µg ml^{-1} chloramphenicol and incubated at 310 K until the OD_{600} reached 0.35. The temperature was shifted to 293 K for 30 min and isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 m*M*. Growth was continued for 5 h, after which the cells were harvested by centrifugation and lysed by sonication in buffer *A* (30 m*M* Tris–HCl pH 7, 150 m*M* NaCl, 10 m*M* imidazole, 5 m*M* β-mercaptoethanol and 1 m*M* PMSF) at 277 K. The lysate was clarified by centrifugation at 12 000*g* for 90 min at 277 K and was loaded onto a 5 ml Ni^{2+}-loaded HiTrap Chelating HP column (GE Healthcare). The SRA protein was eluted with a linear gradient of imidazole from 10 to 300 m*M* in buffer *A*. The SRA fractions were identified by SDS–PAGE, pooled and dialyzed overnight against buffer *A*. The dialyzed protein was digested with recombinant tobacco etch virus (rTEV) protease for 72 h at 277 K at a ratio of 1 mg rTEV protease per 40 mg protein and the reaction mixture was applied onto an Ni^{2+} column. The SRA protein was recovered in the flowthrough fractions, concentrated and loaded onto a size-exclusion chromatography column (Superdex 75 HR 16/50, Amersham Biosciences) equilibrated with 25 m*M* Tris–HCl pH 7, 150 m*M* NaCl, 0.5 m*M* PMSF, 5 m*M* β-mercaptoethanol (hereafter referred to as high-salt buffer).

For the production of selenomethionine (SeMet) labelled protein, the methionine-auxotroph strain B834 (DE3) harbouring both the SRA expression plasmid and the pLysS plasmid was cultured in minimal media M9 containing 17 amino acids, selenomethionine (50 µg ml^{-1}), ampicillin and chloramphenicol (LeMaster & Richards, 1985). The conditions for growing the cells and expressing and purifying the SeMet-labelled protein are identical to those described above for the native protein. Electrospray ionization mass spectrometry of the purified SeMet-labelled SRA protein confirmed the substitution of the three methionine residues by selenomethionine (100% incorporation).

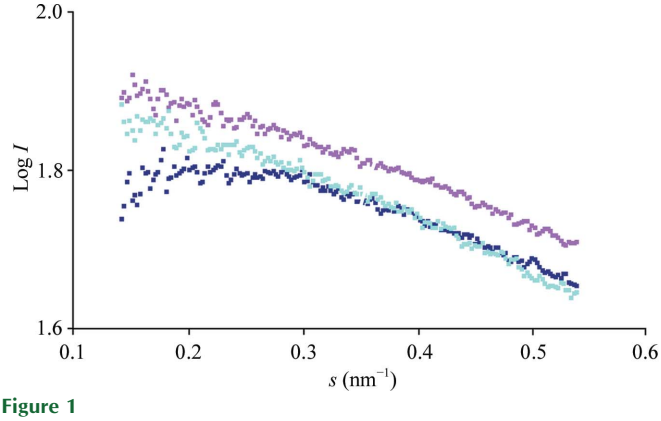

Figure 1
X-ray scattering patterns from the three batches of SRA protein. Experimental scattering is shown from SRA protein in high-salt buffer (cyan squares) and dialyzed against (pink squares) or purified in (dark blue squares) low-salt buffer. The plot displays the logarithm of the scattering intensity as a function of momentum transfer $s = 4\pi\sin(\theta)/\lambda$, where θ is the scattering angle and λ is the X-ray wavelength.

2.2. Dynamic light scattering

Prior to dynamic light-scattering (DLS) experiments, the protein samples were centrifuged for 10 min at 14 000*g*. DLS studies were performed using a DynaPro801 DLS instrument (Protein Solutions Inc.). *DYNAMICS* v.5.25.44 software was used in data collection and analysis. For each experiment, 20 scans were collected at 293 K.

2.3. Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) measurements were performed on DESY beamline X33 (EMBL Hamburg, Germany). The SRA samples were used at a concentration of 3 mg ml^{-1}. Each protein sample was exposed for 120 s at 288 K after the addition of 2 m*M* DTT to eliminate the free radicals formed in solution under X-ray irradiation. Values obtained for buffer solutions without protein were subtracted from the protein data using *PRIMUS* (Konarev *et al.*, 2003). The corrected profiles are shown in Fig. 1.

2.4. Crystallization

The SRA protein was concentrated to 8 mg ml^{-1} in high-salt buffer. Initial crystallization screens were set up with a Mosquito nanodrop robot (Molecular Dimensions) using commercially available kits. Drops consisting of 100 nl protein solution and 100 nl reservoir solution were equilibrated against 80 µl reservoir solution. The initial crystallization condition for SRA was obtained using solution No. 10 of the JCSG screen (Page & Stevens, 2004). After optimization, crystals appeared at 295 K in two or three days in hanging drops composed of 1 µl protein solution at 8 mg ml^{-1} and 1 µl reservoir containing 22–26% unbuffered PEG 3350 solution equilibrated against 1 ml reservoir. Using protein dialyzed against 10 m*M* Tris–HCl pH 7, 20 m*M* NaCl, 5 m*M* $MgCl_2$, 0.1 m*M* TCEP (hereafter referred to as low-salt buffer), crystals appeared in 1 d with the same drop setup and a reservoir containing 14–19% unbuffered PEG 3350 solution. The SeMet-labelled SRA protein behaved similarly to the native protein and crystals of SeMet-labelled protein were

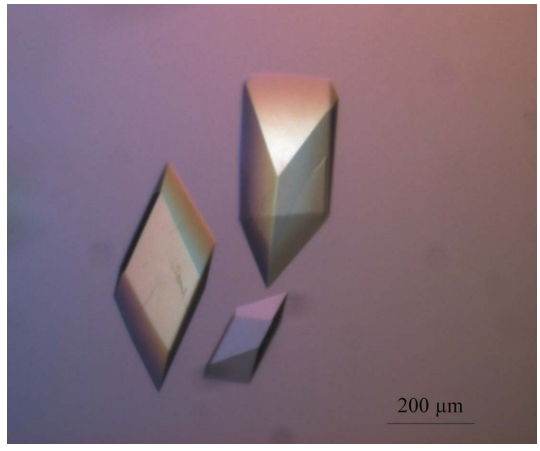

Figure 2
Crystals of SeMet-labelled SRA protein.

obtained under the same conditions. Typical crystal dimensions were 200 × 200 × 400 µm (Fig. 2).

2.5. Data collection and processing

Prior to X-ray analysis, crystals of SeMet-labelled SRA protein were soaked in a cryoprotectant solution composed of 50% reservoir and 50% ethanol by volume for 10 s, which was followed by a quick transfer to paraffin oil prior to flash-cooling in liquid nitrogen.

A complete multiple anomalous dispersion (MAD) data set was collected at 100 K from a single SeMet crystal on the European Synchrotron Radiation Facility (ESRF, Grenoble, France) beamline ID14-EH4 equipped with a Q315 ADSC detector. The fluorescence spectrum recorded from the frozen crystal was used to select the wavelength of the Se *K* absorption edge at the peak (λ = 0.9791 Å, maximum of f''), at the inflection (λ = 0.9795 Å, minimum of f') and at a high-energy remote (λ = 0.9757 Å). 200 frames were recorded at each wavelength using an oscillation range of 0.5° and a crystal-to-detector distance of 320 mm.

The data were processed with *MOSFLM* and scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994). The statistics are summarized in Table 1.

2.6. MAD phasing

The three selenium sites were located with *SOLVE* (Terwilliger & Berendzen, 1999) and were input into a maximum-likelihood heavy-atom parameter refinement using *SHARP* (de La Fortelle & Bricogne, 1997), which was followed by density modification with *SOLOMON* (Abrahams & Leslie, 1996) assuming a solvent content of 50%. A partial model was obtained with *RESOLVE* (Terwilliger, 2000) and was used for further manual building in the experimental electron-density map using *TURBO-FRODO* (Roussel & Cambillau, 1989).

3. Results and discussion

The SRA domain of human UHRF1 was overexpressed in *E. coli* cells with typical yields of about 3 mg per litre of culture for the native protein and 1 mg per litre of culture for the SeMet-labelled protein. The SRA protein was purified by nickel-affinity chromatography with subsequent cleavage of the histidine tag by rTEV. Gel-filtration chromatography as a final purification step provided pure protein as judged by SDS–PAGE with Coomassie staining. Analysis by DLS experiments of the purified SRA protein in high-salt buffer showed a monodisperse solution (polydispersity of 9%) with a hydrodynamic radius of 2.7 nm and a calculated molecular weight of 34 kDa, corresponding to a monomeric species. A low reproducibility rate (<20%) of crystallization was achieved with the protein stored in high-salt buffer. On overnight dialysis of the protein against low-salt buffer, the crystallization rate increased to 70%. However, when the protein was purified in low-salt buffer at the gel-filtration step, the crystallization rate was high but the crystals dissolved within 2 d. SAXS experiments indicated monodispersity of the SRA solutions when the protein was in high-salt buffer or subsequently dialyzed against low-salt buffer but the occurrence of repulsive interactions when the protein was purified in low-salt buffer (Fig. 1). Numerous and stable protein crystals that were suitable for diffraction experiments were thus obtained with the SeMet-labelled protein purified in high-salt buffer and dialyzed against low-salt buffer.

The crystals belong to the trigonal space group $P3_221$, with unit-cell parameters $a = b = 53.78$, $c = 162.05$ Å. They contain one molecule in the asymmetric unit, with a corresponding Matthews coefficient (Matthews, 1968) of 2.54 Å^3 Da^{-1} and a solvent content of 52%. A MAD data set collected to 2.2 Å resolution allowed the structure determination of the human SRA domain. 73% of the atoms were built in the experimental electron-density map, attesting to the good quality of the MAD phases. A detailed description of the refined structure will be published elsewhere (PDB code 2pb7).

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Data set	Peak	Inflection	Remote
Beamline	ID14-EH4		
Wavelength (Å)	0.9791	0.9795	0.9757
Resolution (Å)	50–2.20 (2.32–2.20)		
Space group	$P3_221$		
Unit-cell parameters			
$a = b$ (Å)	53.78		
c (Å)	162.05		
Observations	81284	81449	81301
Unique reflections	14534	14551	14529
Completeness (%)	99.9 (99.9)	99.9 (100.0)	99.8 (99.9)
Redundancy	5.6 (5.8)	5.6 (5.8)	5.6 (5.8)
Anomalous redundancy	3.0 (3.0)	3.0 (3.0)	3.0 (3.0)
$R_{\rm merge}$†	0.085 (0.296)	0.081 (0.295)	0.080 (0.289)
$R_{\rm p.i.m.}$‡	0.047 (0.146)	0.041 (0.147)	0.043 (0.145)
$R_{\rm r.i.m.}$§	0.111 (0.358)	0.098 (0.354)	0.100 (0.348)
$R_{\rm ano}$¶	0.059 (0.153)	0.042 (0.142)	0.050 (0.146)
$I/\sigma(I)$	15.6 (5.1)	16.6 (5.1)	16.3 (5.2)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observed intensity of a measured reflection hkl and $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$. ‡ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where N is the redundancy. § $R_{\rm r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl))$, where N is the redundancy. ¶ $R_{\rm ano} = \sum_h |\langle I(h+)\rangle - \langle I(h-)\rangle| / \sum_h [\langle I(h+)\rangle + \langle I(h-)\rangle]$, where $\langle I(h+)\rangle$ and $\langle I(h-)\rangle$ correspond to the average intensities of each Friedel pair for reflection h.

We thank the scientific staff of ESRF beamline ID14-EH4 (Grenoble, France) and DESY beamline X33 (Hamburg, Germany) for providing excellent data-collection facilities, Dino Moras for his contribution and discussions, the Structural Biology and Genomics Platform (especially Edouard Troesch) for cloning, Guillaume Bec for technical assistance with the Mosquito nanodrop robot and Hélène Nierengarten for mass-spectrometric analysis. This work was supported by the Region Alsace, the European Commission as SPINE2-Complexes (contract No. LSHG-CT-2006-031220) and 'Cancéropôle Grand-Est'.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Achour, M., Jacq, X., Rondé, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A., Schini-Kerth, V. & Bronner, C. (2008). *Oncogene*, **27**, 2187–2197.
Bostick, M., Kim, J., Esteve, P., Clark, A., Pradhan, S. & Jacobsen, S. E. (2007). *Science*, **317**, 1760–1764.
Busso, D., Delagoutte-Busso, B. & Moras, D. (2005). *Anal. Biochem.* **343**, 313–321.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruz, X. de la, Lois, S., Sanchez-Molina, S. & Martinez-Balbas, M. (2005). *BioEssays*, **27**, 164–175.
Jenuwein, T. & Allis, C. D. (2001). *Science*, **293**, 1074–1080.
Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J. & Jacobsen, S. E. (2007). *Curr. Biol.* **17**, 379–384.
Karagianni, P., Amazit, L., Qin, J. & Wong, J. (2008). *Mol. Cell. Biol.* **28**, 705–717.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

LeMaster, D. & Richards, F. (1985). *Biochemistry*, **24**, 7263–7268.
Matthews, B. (1968). *J. Mol. Biol.* **33**, 491–497.
Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A. C., Dorn, R., Jenuwein, T. & Reuter, G. (2005). *EMBO J.* **24**, 1418–1429.
Page, R. & Stevens, R. (2004). *Methods*, **34**, 373–389.
Papait, R., Pistore, C., Grazini, U., Babbio, F., Cogliati, S., Pecoraro, D., Brino, L., Morand, A., Dechampesme, A., Spada, F., Leonhardt, H., McBlane, F., Oudet, P. & Bonapace, I. M. (2008). *Mol. Biol. Cell*, **19**, 3554–3563.
Papait, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L. & Bonapace, I. M. (2007). *Mol. Biol. Cell*, **18**, 1098–1106.
Roussel, A. & Cambillau, C. (1989). *TURBO-FRODO*. Silicon Graphics, Mountain View, California, USA.
Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M. & Koseki, H. (2007). *Nature (London)*, **450**, 908–913.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Unoki, M., Nishidate, T. & Nakamura, Y. (2004). *Oncogene*, **23**, 7601–7610.
Woo, H. R., Pontes, O., Pikaard, C. S. & Richards, E. J. (2007). *Genes Dev.* **21**, 267–277.