Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Definition of domain boundaries and crystallization of the SMN Tudor domain

Remco Sprangers,[a] Philipp Selenko,[a] Michael Sattler,[a] Irmgard Sinning[b] and Matthew R. Groves[b]*

[a]Structural and Computational Biology, EMBL, Meyerhofstrasse 1, D-69012, Heidelberg, Germany, and [b]Biochemie-Zentrum (BZH), University of Heidelberg, Im Neuenheimer Feld 328, D-69120, Heidelberg, Germany

Correspondence e-mail: matthew.groves@bzh.uni-heidelberg.de

Received 2 September 2002
Accepted 20 November 2002

This paper is dedicated to the memory of Dr V. Dhanaraj.

Spinal muscular atropy (SMA) is the major genetic disease leading to childhood mortality and is caused by mutations in or deletions of the *smn1* gene. The human survival of motor neurons (SMN) protein encoded by this gene plays an important role in the assembly of snRNPs (small nuclear ribonucleoprotein complexes) *via* binding to the spliceosomal Sm proteins. The tails of these Sm proteins contain symmetrically dimethylated arginines that are recognized by the central SMN Tudor domain. To gain insight in the molecular basis of this specific interaction, the SMN Tudor domain has been crystallized. The rapid crystallization of the protein and the high stability of the crystals is facilitated by redefinition of domain boundaries based on NMR relaxation experiments and the previously determined solution structure. The crystals diffract to high resolution (1.8 Å) and a complete data set has been collected from a hexagonal crystal form ($P6_1/P6_5$), with unit-cell parameters $a = b = 27.65$, $c = 110.30$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Crystal soaks and co-crystallization with peptides derived from the Sm protein tails have been initiated. Molecular replacement with the NMR coordinates is under way.

1. Introduction

The neuromuscular disease spinal muscular atrophy (SMA) is characterized by loss of the motor neurons of the spinal cord. SMA results from mutations or deletions in the *smn1* gene, leading to reduced levels of full-length functional SMN (survival of motor neurons) protein (Brzustowicz *et al.*, 1990; Lefebvre *et al.*, 1995) and as a consequence to muscle weakness and atrophy. With a carrier frequency of 1 in 50, the disease occurs in 1 in 10 000 live births and is the major genetic cause of childhood mortality.

The SMN protein is found in all metazoan cells and is part of a large protein complex. In the nucleus, the SMN protein plays a role in pre-mRNA metabolism. In the cytoplasm, the SMN protein plays a crucial role in the assembly of the spliceosomal uridine-rich small nuclear ribonucleoprotein complexes (U snRNP; Liu *et al.*, 1997). Those complexes consist of the uridine-rich snRNAs and seven spliceosomal Sm proteins, which are thought to form a heptameric ring around the U snRNA (Kambach *et al.*, 1999). In addition to two Sm sequence motifs, some of these Sm proteins contain an additional C-terminal tail that is rich in arginine and glycine residues. During snRNP assembly, the SMN Tudor domain binds to the tails of the spliceosomal Sm core proteins (Bühler *et al.*, 1999) and antibodies against the SMN Tudor domain prevent splicing *in vivo*. The Tudor domain harbouring the E134K mutation loses the ability to bind to the RG (arginine/glycine) rich Sm tails and the mutation leads to a severe form of SMA.

Previously, we determined the three-dimensional structure of the Tudor domain using heteronuclear NMR (nuclear magnetic resonance) spectroscopy (Selenko *et al.*, 2001). The structure consists of a long curved β-sheet forming a β-barrel, similar to the fold of the Sm proteins. In previous studies, we also showed that the C-terminal RG-rich tails of the Sm proteins bind specifically to a hydrophobic patch on the surface of the SMN Tudor domain. *In vivo*, the RG-rich tails of the Sm proteins are post-translationally modified and contain symmetrically dimethylated arginines (sDMAs; Brahms *et al.*, 2000). This modification has been shown to increase the binding of Sm tails to the SMN Tudor domain and is suggested to be a regulation mechanism during assembly of the spliceosome. Interestingly, other proteins that bind to the SMN complex also contain RG-rich domains that can potentially be methylated. Those domains are important for the association to the SMN complex (Paushkin *et al.*, 2002); the SMN Tudor domain might thus also play a role in regulation of these protein–protein interactions.

In order to obtain structural information on the interaction between the Tudor domain and the Sm tails containing the sDMAs, we

followed two complementary approaches. The first involves the use of heteronuclear magnetic resonance spectroscopy to define the residues of the Tudor domain that are involved in the interaction and to define the exact boundaries of the protein domain. The second method involves X-ray crystallography to obtain high-resolution structural data of the complex, which because of the lack of nuclear Overhauser effects (NOEs) between the protein and peptide could so far not be obtained using NMR spectroscopy. Here, we report the crystallization of the SMN Tudor domain alone. Soaks and co-crystallizations with peptides derived from the Sm tails, containing sDMAs, have been initiated. For the uncomplexed protein, a complete data set has been collected to a resolution of 1.8 Å at ESRF beamline ID29.

2. Materials and methods

2.1. Expression, purification and crystallization

The Tudor domain of SMN was PCR-amplified from a human cDNA library (Genbank accession No. Q16637). Flexible regions C-terminal and N-terminal to the SMN Tudor domain were identified based on the previously determined NMR structure and NMR relaxation experiments recorded on a Bruker DRX 600 NMR spectrometer (Farrow *et al.*, 1994). Human SMN cDNA was designed accordingly and inserted into the *Nco*I/*Kpn*I sites of a modified pET24d expression vector (Novagen) containing an N-terminal His_6-GST tag followed by a TEV protease-cleavage site. The expression clone was confirmed by DNA sequencing. The 66-residue recombinant protein used for the structural studies comprised residues 82–147 of SMN protein plus three additional residues from the TEV cleavage site (GAM; Fig. 1).

The Tudor domain was overexpressed in *Escherichia coli* BL21(DE3) by induction with 1 m*M* IPTG at 298 K for 4 h. The cell lysate was fractionated over a nickel column and the bound protein was eluted with 0.3 *M* imidazole. The N-terminal His_6-GST tag was cleaved off with N-His_6-TEV protease overnight at room temperature. After changing to a buffer containing a low imidazole concentration, the TEV protease and the uncleaved protein were removed over a nickel column. The flow-through was loaded onto a Resource Q column and the bound Tudor domain was eluted from the column by a NaCl gradient to 1 *M*. Subsequent gel filtration (in 300 m*M* NaCl, 20 m*M* Tris–HCl and 2 m*M* DTT pH 8.0) was performed on a Sephadex HiPrep 16/60 column (Pharmacia) and resulted in SMN Tudor domain purified to apparent homogeneity. The expression products were checked by mass spectrometry [7733.5 (1) Da; expected value, 7733.7 Da].

2.2. Crystallization

The protein was concentrated using Amicon Centricon 3 kDa filters to a concentration of 1.5 m*M* as estimated from the absorption at 280 nm (ε_0 = 15220 M^{-1} cm^{-1}). The hanging-drop vapour-diffusion technique was used to screen for suitable crystallization conditions using Wizard Screens I and II (Emerald Biosciences). Drops were prepared on siliconized cover slips by combining equal volumes (2 µl each) of reservoir solution and protein solution and were equilibrated against 500 µl reservoir solution. Conditions producing crystals from the initial screens were refined to produce crystals suitable for X-ray diffraction analysis.

2.3. Data collection

After transferring to a cryoprotectant solution [reservoir solution with an additional 30%(*v*/*v*) glycerol], crystals were picked up using a fibre loop and flash-frozen in a stream of nitrogen gas at 100 K. In-house diffraction data were collected from a single crystal on a MAR345 (MAR Research) image-plate detector using Cu $K\alpha$ radiation from a rotating-anode X-ray generator operating at 50 kV and 100 mA. A high-resolution data set was collected on ESRF beamline ID29 using an ADSC detector. The programs *MOSFLM* (Leslie, 1992), *SCALA* (Collaborative Computational Project, Number 4, 1994) and *CNS* (Brünger *et al.*, 1998) were used for initial data processing and analysis.

3. Results and discussion

The NMR-optimized crystallization construct was designed based on heteronuclear {^{1}H}-^{15}N NOE experiments (Farrow *et al.*, 1994), which give information about the mobility of the protein backbone on a residue-by-residue basis on a picosecond to nanosecond timescale. No attempts to crystallize the initial NMR construct were performed, as around 30% of the NMR construct was disordered or unfolded (Fig. 1).

Two conditions from the initial screen (2.0 *M* ammonium sulfate, 0.1 *M* sodium citrate pH 5.5; 2.0 *M* ammonium sulfate, sodium phosphate/citrate pH 4.5) produced microcrystals from which further screens were designed. Hexagonal-shaped crystals large enough for X-ray analysis were obtained in 2.5 *M* ammonium sulfate and 0.1 *M* sodium citrate pH 4.5. These crystals appeared after 1–5 h, were suitable for analysis after about 12 h (0.4 × 0.2 × 0.2 mm; see Fig. 2) and were stable for at

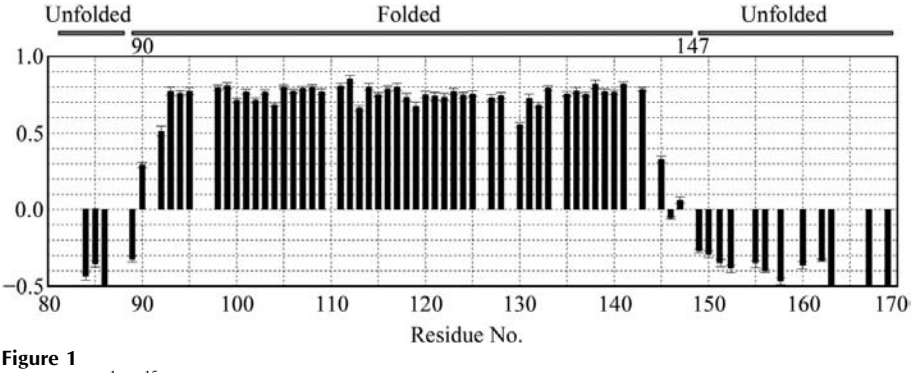

Figure 1
Plot of the {^{1}H}-^{15}N heteronuclear NOE value *versus* residue number. Positive and negative values of the heteronuclear NOE indicate folded and unfolded regions of the protein, respectively, as indicated at the top. The 66-residue recombinant protein used for the structural studies comprised residues 82–147 of SMN protein plus three additional residues from the TEV cleavage site (GAM).

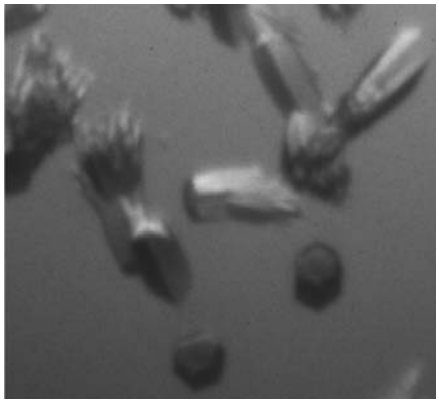

Figure 2
Crystals of the SMN Tudor domain.

Table 1
X-ray data-collection statistics of the Tudor domain on ESRF beamline ID29.

Values in parentheses correspond to the highest resolution shell.

Space group	Hexagonal ($P6_1/P6_5$)
Unit-cell parameters	
$a = b$ (Å)	27.65
c (Å)	110.3
$\alpha = \beta$ (°)	90
γ (°)	120
Wavelength (Å)	0.93
Temperature (K)	100
Exposure time (s)	5
Oscillation range per frame (°)	2
Resolution (Å)	36.8–1.80 (1.90–1.80)
Observed reflections	51260 (7493)
Unique reflections	4489 (652)
$R_{\rm merge}$ (%)	11.7 (48.5)
Completeness (%)	100 (100)
$I/\sigma(I)$	3.4 (1.0)
Unique reflections $I > 3\sigma(I)$ (%)	56.4
Multiplicity	11.4 (11.5)

least several months. The very fast crystallization of the protein and the long-term stability of the crystals is likely to be aided by the exact definition of the protein-domain boundaries as derived from NMR experiments. Diffraction from these crystals was measured in-house to 3.0 Å. At the ESRF beamline ID29, a complete high-resolution data set was recorded at a crystal-to-detector distance of 200 mm and an oscillation range of 180° in 2.0° steps to a resolution of 1.8 Å.

Autoindexing yielded unit-cell parameters $a = b = 27.65$, $c = 110.30$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ in the hexagonal point group $P6$ (data-collection parameters are given in Table 1).

Subsequent processing of the data indicated the presence of a screw axis ($P6_1/P6_5$). Using a molecular weight of 7.6 kDa and one molecule per asymmetric unit, the predicted solvent content for one molecule in the asymmetric unit is 11% (Matthews, 1968).

Efforts are now under way to use the NMR structure as a molecular-replacement search model. Co-crystallization and crystal soaks with small peptides derived from the Sm tails and containing symmetrically dimethylated arginines have been initiated. Furthermore, crystal soaks with a single symmetrically dimethylated arginine have been started. The small size of this single amino acid increases the possibility of the ligand entering the crystal lattice during the soaking experiment. A three-dimensional structure of the complex of the SMN Tudor domain with methylated arginines will reveal the molecular basis of the specific recognition of methylated arginines and the defect underlying SMA arising from the E134K mutation.

The authors are grateful to Gunter Stier for support and assistance with the protein expression and purification. The authors thank Elena Conti for initial diffraction experiments and the ESRF ID29 beamline staff for assistance during the data collection. This work has been supported by a DFG grant to MS and the authors would also like to acknowledge the support of the EMBL.

References

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L. & Luhrmann, R. (2000). *J. Biol. Chem.* **275**, 17122–17129.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppart, M., Ziter, F., Wood, D., Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L. & Gilliam, T. C. (1990). *Nature (London)*, **344**, 540–541.
Bühler, D., Raker, V., Lührmann, R. & Fischer, U. (1999). *Hum. Mol. Genet.* **8**, 2351–2357.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). *Biochemistry*, **33**, 5984–6003.
Kambach, C., Walke, S., Young, R., Avis, J. M., de La Fortelle, E., Raker, V. A., Luhrmann, R., Li, J. & Nagai, K. (1999). *Cell*, **96**, 375–387.
Lefebvre, S., Bürglen, L., Reboullet, S., Cermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A. & Melki, J. (1995). *Cell*, **80**, 155–165.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACB Newsl. Protein Crystallogr.* **26**.
Liu, Q., Fischer, U., Wang, F. & Dreyfuss, G. (1997). *Cell*, **90**, 1013–1021.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Paushkin, S., Gubitz, A. K., Massenet, S. & Dreyfuss, G. (2002). *Curr. Opin. Cell. Biol.* **14**, 305–312.
Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U. & Sattler, M. (2001). *Nature Struct. Biol.* **8**, 27–31.