Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, crystallization and preliminary X-ray characterization of the *Drosophila* transcription factor Doublesex

James Bayrer, Zhuli Wan, Biaoru Li and Michael A. Weiss*

Case Western Reserve University, Cleveland, Ohio, USA

Correspondence e-mail: weiss@biocserver.bioc.cwru.edu

Received 4 May 2004
Accepted 19 May 2004

Doublesex (DSX) is a transcription factor in *Drosophila melanogaster* that regulates somatic sexual differentiation. Homologues have been found in diverse metazoans, suggesting a conserved role in development. Here, the expression, purification and crystallization of a novel 63-amino-acid dimerization domain and selenomethionine analogue are described. Native crystals belong to space group $P2_12_12_1$ and diffract to a resolution of 1.6 Å with synchrotron radiation. Selenomethionine-containing crystals provide anomalous dispersion for phasing.

1. Introduction

Doublesex (DSX) is a downstream component of the well described sex-determining hierarchy of *Drosophila melanogaster*. In this genus sex is determined by the ratio of X chromosomes to autosomes. When the ratio is 2:2 (XX genotype), the *sex lethal* gene (encoding an RNA splicing factor) is activated, leading to the expression of Transformer (TRA), a second RNA-splicing factor. TRA, together with TRA2, regulates the female splicing isoform of DSX (DSX^F). When the ratio of X chromosomes to autosomes is 1:2 (XO genotype), *sex lethal* is not expressed and the male isoform of DSX (DSX^M) is produced by default. Messenger RNAs encoding DSX^M and DSX^F contain the same first three exons; the C-terminal region of DSX^F is encoded by exon four, whereas the C-terminal region of DSX^M is encoded by exons five and six (Burtis & Baker, 1989). Thus, as a consequence of sex-specific splicing, male and female isoforms are identical for the first 397 residues but differ for the remainder of the proteins.

The N-terminal domain of DSX (residues 35–105) contains a DNA-binding motif, designated the DM domain (Doublesex and MAB-3 domain; Raymond *et al.*, 1998). As the acronym implies, this domain is also found in *Caenorhabditis elegans* transcription factor MAB-3, which is required for male somatic differentiation. Indeed, studies of transgenic worms have demonstrated that the male (but not female) isoform of DSX can rescue male features in *C. elegans* in the absence of MAB-3 (Raymond *et al.*, 1998). Recently, the DM domain's role in sexual differentiation has been shown to be conserved among metazoans. In humans, for example, deletion of three contiguous DM-containing genes on chromosome 9 is associated with XY sex reversal and gonadoblastoma (the 9p syndrome; Ottolenghi *et al.*, 2000). The structure of the DM domain, determined by NMR spectroscopy, contains a novel zinc module that binds in the minor groove of DNA (Zhu *et al.*, 2000).

The C-terminal domain of DSX (residues 350–427 in DSX^F and 350–549 in DSX^M) contains both sex-specific and non-sex-specific residues. This domain is responsible for oligomerization and presumably for the differential transcriptional regulation effected by male and female isoforms (Erdman *et al.*, 1996). A mutation in XX (genetically female) fruit flies at position 398 (G398D) results in an intersex phenotype. This mutation impairs dimerization *in vitro*, implying that dimerization is required for the function of DSX^F *in vivo* (Erdman *et al.*, 1996). Homologues of the dimerization domain are not recognizable outside of insects. Given the broad conservation of the DM domain, however, it is possible that the three-dimensional structure of the dimerization domain will be shared by proteins with limited sequence similarity. In this report we describe the expression, purification and crystallization of the DSX^F dimerization domain, its labeling with selenomethionine, and preliminary crystallographic X-ray characterization.

2. Materials and methods

2.1. Protein expression and purification

Coding sequences corresponding to the C-terminal domain of DSX^F (residues 350–427) and a truncated analogue (residues 350–412) were ligated into expression plasmid pMW127 (Hinck *et al.*, 1993; Ukiyama *et al.*, 2001). The truncated analogue was designed based on prior yeast two-hybrid studies showing that residues 413–427 are not required for dimerization (Erdman *et al.*, 1996); sequence analysis suggests that this tail is unstructured. The domains were expressed in *Escherichia coli* strain BL21(DE3)pLysS (Invitrogen, Carlsbad, California, USA) as

© 2004 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Data-collection and analysis statistics.

Values in parentheses are for the highest resolution shell.

	Native	SeMet
Wavelength (Å)	0.9000	0.9788
Unit-cell parameters (Å)	a = 39.8, b = 46.6, c = 59.8	a = 39.4, b = 46.8, c = 58.6
Resolution range (Å)	19–1.6 (1.66–1.60)	32.7–1.8 (1.86–1.80)
No. measured reflections	164648	145768
No. unique reflections	15159 (1467)	19414 (1905)
Redundancy	10.9	7.5
Completeness (%)	99.1 (96.9)	100 (100)
$R_{\rm merge}$ (%)	3.8 (0.182)	8.7 (0.456)
Average $I/\sigma(I)$	62.3 (11)	24.0 (3.4)

thrombin-cleavable fusion proteins with an N-terminal His_6 tag under the transcriptional control of isopropyl-β-D-thiogalactoside (IPTG; Roche, Indianapolis, Indiana, USA). Essentially identical protocols were employed for protein purification. In each case, a 120 ml starter culture containing 2YT medium and antibiotics was grown overnight with shaking at 310 K and subsequently inoculated into 4.5 l of 2YT containing ampicillin (AMP; Roche, Indianapolis, Indiana, USA) and chloramphenicol (CA; Boehringer Mannheim GmbH, Mannheim, Germany). Cultures were grown to an optical density at 600 nm (OD_{600}) of ∼0.6 and then induced with 0.5 m*M* IPTG. Cells were grown to an OD_{600} of ∼1.0 and harvested by centrifugation at 3500*g* for 15 min at 277 K. The cell pellet was harvested, resuspended in extraction buffer (20 m*M* imidazole, 20 m*M* Tris–HCl pH 7.0, 200 m*M* NaCl) with 0.75 mg ml^{-1} lysozyme and 0.174 mg ml^{-1} phenylmethylsulfonyl fluoride (PMSF; Sigma, St Louis, Missouri, USA) and incubated for 20 min at room temperature. The mixture was then exposed to three freeze–thaw cycles followed by sonication on ice. The lysate was centrifuged at 17 400*g* for 1 h at 277 K. The supernatant was reserved, and the pellet was resuspended in extraction buffer, sonicated and centrifuged again. The supernatants were combined, incubated with Co^{2+}-affinity resin (BD Biosciences, Palo Alto, California, USA), and applied to an FPLC column at 277 K. The column was washed with two to three column volumes of extraction buffer. The fusion protein was eluted from the column with elution buffer (250 m*M* imidazole, 20 m*M* Tris–HCl pH 7.0, 200 m*M* NaCl) and dialyzed into cleavage buffer (5 m*M* imidazole, 20 m*M* Tris–HCl pH 7.9, 200 m*M* NaCl) overnight at 277 K. Thrombin cleavage was accomplished at room temperature with *ca* one unit of thrombin (Sigma, St Louis, Missouri, USA) per milligram of fusion protein until completion as assayed by SDS–PAGE, usually ∼4 h. The mixture was applied to a Co^{2+}-affinity column, and the flowthrough containing the protein of interest was collected. Eluted fractions containing the domain were concentrated using an Amicon ultrafiltration system with a YM1000 filter (Millipore, Billerica, Massachusetts, USA). The protein was purified to near-homogeneity by gel filtration (Superdex 75; Amersham Pharmacia, Uppsala, Sweden) in the case of the truncated domain and by anion exchange (DEAE-5PW; Toso Haas, GmbH, Stuttgart, Germany) in the case of the intact domain.

Selenomethionine-labeled protein was expressed in B834(DE3)pLysS competent cells (Novagen, Madison, Wisconsin, USA) grown in M9 minimal medium supplemented with all amino acids except methionine at 40 mg l^{-1} and selenomethionine (Sigma, St Louis, Missouri, USA) at 50 mg l^{-1}. A 140 ml starter culture containing 135 ml supplemented M9 media, and 5 ml 2YT media and antibiotics was grown overnight at 310 K. The starter culture (120 ml) was inoculated into 4.5 l supplemented M9 medium together with 6 ml 2YT medium to facilitate cell growth. Induction, harvest and purification were as described for the native domains.

2.2. Crystallization

The purified native fragment (residues 350–412) was concentrated in an Amicon ultrafiltration cell using YM1000 filters to ∼10 mg ml^{-1} in GPC buffer (20 m*M* Tris–HCl pH 7.4, 200 m*M* NaCl). Crystallization trials were performed with Crystal Screen I and II and PEG/Ion Screen using the hanging-drop vapor-diffusion method (Hampton Research, Laguna Niguel, California, USA). Drops were formed by mixing 1–4 µl protein stock with 1–4 µl reservoir solution. Crystallization drops were allowed to equilibrate at either room temperature or 277 K for 1–6 weeks over 1 ml of reservoir solution. Conditions initially producing crystals were optimized by varying the precipitant and protein concentrations. Optimal crystallization conditions for the dimerization domain were determined to be 1.85 *M* ammonium sulfate and 7% 2-propanol with a protein stock concentration of 12 mg ml^{-1}. Crystallization conditions for the SeMet-containing analogue were identical to those given above. Crystals were not obtained for the intact C-terminal domain (residues 350–427) under any of the conditions tested.

2.3. Data collection and processing

Native X-ray diffraction data were collected under cryoconditions on beamline 14-ID-B at Argonne National Laboratories. Prior to data collection, crystals were washed in a solution containing reservoir solution and 15%(*v*/*v*) glycerol, mounted in nylon loops and flash-frozen in liquid nitrogen. Data collection was at 0.9000 Å at a crystal-to-detector distance of 110 mm and an oscillation range of 1°. Single-wavelength anomalous dispersion (SAD) data were collected under cryoconditions on beamline X9B at Brookhaven National Laboratories. Data reduction was performed using *DENZO* (native) or *HKL*2000 (SAD). Scaling was performed with *SCALEPACK* (native) or *HKL*2000 (SAD) (Otwinowski & Minor, 1997). Substructure determination and phasing were accomplished using the *SHELX* suite of programs (Sheldrick, 1990). Data-collection statistics are given in Table 1.

3. Results and discussion

3.1. Expression and purification

Protein expression was tested with 2 ml cultures grown from single colonies. Those cultures showing greatest overexpression, as determined by SDS–PAGE, were selected for large-scale protein expression. After elution from the first cobalt column, a band of slightly less than the molecular weight of the fusion protein (36.5 kDa for the dimerization domain, 34.5 kDa for the analogue) was in each case observed by SDS–PAGE. DSX has been noted to migrate abnormally on SDS–PAGE gels (Burtis *et al.*, 1991; Cho & Wensink, 1996). The purified proteins were assayed by mass spectrometry and found to correspond to the predicted molecular weights of the domain plus Gly-Ser derived from the thrombin-cleavage site (9.402 kDa for the dimerization domain, 7.737 kDa for the truncated analog). It was noted that protein purified by reverse-phase high-performance liquid chromatography (HPLC) was unusually susceptible to oxidation and aggregation. Purification of the domains by gel filtration or ion exchange under non-denaturing conditions resulted in each case in a stock solution stable at room temperature and amenable to crystallization trials. Selenomethionine protein was purified similarly. Protein yields were typically 8 mg per litre of culture under native conditions and 6 mg per litre of culture for incorporation of selenomethionine.

Figure 1
Crystals of DSX^{F} WT (residues 350–412) grown in 1.9 *M* ammonium sulfate and 7% 2-propanol.

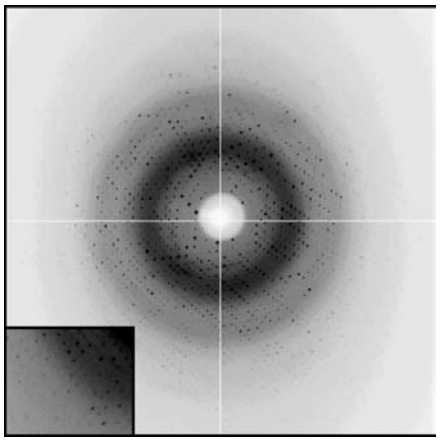

Figure 2
X-ray diffraction pattern of native DSX^{F} WT 350–412 collected at APS. Exposure time 2 s, distance 110 mm, oscillation range 1.0°. An ADSC Q4 CCD detector was used to record the image. The frame-edge close-up is 1.6 Å.

3.2. Crystallization

Initial trials of DSX^{F} WT 350–427 failed to yield crystals. NMR and circular-dichroism (CD) studies suggest that this domain consists of a well ordered core dimer with a disordered tail (unpublished results). Reasoning that this tail might be impairing crystallization, we pursued crystallization of a dimeric fragment lacking this tail (residues 350–412). Trials were initially started with an 8.5 mg ml^{-1} protein stock solution. Crystals were observed after three weeks in 2.0 *M* ammonium sulfate and 5% 2-propanol. Conditions were optimized to 1.85 *M* ammonium sulfate and 7% 2-propanol with a 12 mg ml^{-1} protein stock solution. Under these conditions, crystals (Fig. 1) began forming within 3 d, with the most useful crystals forming within one week at room temperature. Crystal dimensions ranged from 0.1 × 0.1 × 0.1 mm to 0.5 × 0.5 × 0.3 mm, the best diffracting crystals being 0.3 × 0.3 × 0.2 mm in size. Selenomethionine crystals were grown under the same conditions as the native crystals. The morphology of the selenomethionine crystals varied slightly from that of the native crystals. Whereas both were roughly square-shaped, the selenomethionine crystals were generally thinner with an 'X' across the face. As indicated by the X-ray diffraction pattern, this 'X' did not appear to arise from satellite crystal formation.

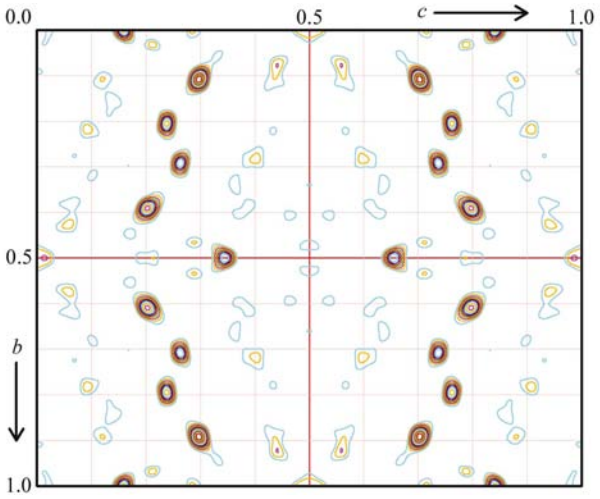

Figure 3
Harker section from anomalous scattering Patterson map showing selenium positions.

3.3. Data collection and preliminary X-ray characterization

Native crystals diffracted to 1.6 Å resolution on beamline 14-ID-B at Argonne National Laboratories (Fig. 2). The crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 39.773, b = 46.623, c = 59.771 Å. Assuming one dimer per asymmetric unit (15.4 kDa), the Matthews coefficient is 1.79 Å^3 Da^{-1} with a solvent content of 30%. To obtain experimental phases, an SeMet-derivative crystal was measured at the Se peak (λ = 0.9788 Å). Data to 1.8 Å resolution were collected on beamline X9B at Brookhaven National Laboratory. The crystal parameters were similar to those of native crystals (space group $P2_12_12_1$; unit-cell parameters a = 39.414, b = 46.759, c = 58.596 Å). Four selenium sites were located in the asymmetric unit (Fig. 3). Although the intact dimerization domain (residues 350–427) appears to be refractory to crystallization, it is possible that NMR spectroscopy may enable extension of the pending structure of the 350–412 domain to include the missing tail.

We thank B. Baker (Stanford University) for helpful discussion, N. Phillips (CWRU) for aid in protein purification, and G. Reddy (University of Chicago) for mass spectrometry. This work is a contribution from the Cleveland Center for Structural Biology. James Bayrer is a predoctoral fellow of the Medical Scientist Training Program at the Case School of Medicine (grant number T32 GM07250).

References

Burtis, K. C. & Baker, B. S. (1989). *Cell*, **56**, 997–1010.
Burtis, K. C., Coschigano, K. T., Baker, B. S. & Wensink, P. C. (1991). *EMBO J.* **10**, 2577–2582.
Cho, S. & Wensink, P. C. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 2043–2047.
Erdman, S. E., Chen, H. J. & Burtis, K. C. (1996). *Genetics*, **144**, 1639–1652.
Hinck, A. P., Walkenhorst, W. F., Westler, W. M., Choe, S. & Markley, J. L. (1993). *Protein Eng.* **6**, 221–227.
Ottolenghi, C., Vietia, R., Quintana-Murci, L., Torchard, D., Scapoli, L., Souleyrear-Therville, N., Beckmann, J., Fellous, M. & McElreavey, K. (2000). *Genomics*, **64**, 170–178.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J. & Zarkower, D. (1998). *Nature* (*London*), **391**, 691–695.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Ukiyama, E., Jancso-Radek, A., Li, B., Milos, L., Zhang, W., Phillips, N. B., Morikawa, N., King, C. Y., Chan, G., Haqq, C. M., Radek, J. T., Poulat, F., Donahoe, P. K. & Weiss, M. A. (2001). *Mol. Endocrinol.* **15**, 363–377.
Zhu, L., Wilken, J., Phillips, N. B., Narendra, U., Chan, G., Stratton, S. M., Kent, S. B. & Weiss, M. A. (2000). *Genes Dev.* **14**, 1750–1764.