Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural studies of the Msx-1 homeodomain–DNA complex I

Stacy DeWees and James H. Geiger*

Michigan State University Chemistry Department, East Lansing, MI 48824, USA

Correspondence e-mail: geiger@cem.msu.edu

Received 23 April 1999
Accepted 20 September 1999

Crystals of the Msx-1 homeodomain–DNA complex have been obtained by hanging-drop vapor diffusion at 293 K in 12% PEG 4000 and 0.1 *M* sodium acetate pH 4.6. The homeodomain consists of 60 amino acids and is the DNA-binding domain. The DNA in the complex was 16 base pairs with the sequence 5′-TGTCACTAATT-GAAGG-3′, containing an overhang T at each end. The crystals diffract to 2.15 Å (99.8% completeness) using cryogenic (123 K) conditions. The crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 33.66, b = 60.96, c = 83.37 Å. The structure will illuminate the details of Msx-1–DNA binding specificity and clarify its role in transcriptional regulation. Mutations in Msx-1 cause craniofacial deformities in mice.

1. Introduction

The study of the molecular processes that regulate mouse embryonic development led to the identification of numerous genes whose protein products control gene expression during embryogenesis. These are transcriptional regulatory proteins that establish and maintain the appropriate patterns of spatial and temporal gene expression (Catron *et al.*, 1993). These genes share the conserved homeobox that encodes the homeodomain. The homeodomain is thought to be a scaffold for protein–protein interactions *in vivo*. The common feature among transcriptional regulatory proteins is that they function to transduce cellular signaling events to changes in gene expression. The murine gene *hox*7.1 (*Msx*-1) is a vertebrate subclass of the Msx gene family and is expressed throughout the developing neural tube (Catron *et al.*, 1993). Msx-1 in the mouse has shown expression in many organs and tissues. It also appears to be involved in determining the specific shapes of bones along with bone morphogenic proteins such as osteocalcin (Davidson, 1995). Mutations in Msx-1 can cause craniofacial deformities in mice. The consensus DNA site (ACTAATTG) and the homeodomain of Msx-1 were used to analyze the transcriptional properties of Msx-1. The precise DNA sequence of these sites leads to functional specificity of the homeodomain proteins *in vivo*. Differences found in N-terminal sequences between various homeodomains translate into different binding preferences, which then contribute to different biological functions. The N-terminal arm plays an important role in this process. Msx-1 acts as a transcriptional repressor of TATA-containing and TATA-less promoters both *in vivo* and *in vitro* (Catron *et al.*, 1993, 1995). The mechanism of this repression appears to involve direct interactions with TBP, as point mutations deficient in TBP binding are also deficient in repression. DNA-binding specificity together with the association with other transcription factors could account for functional specificity. Msx-1 forms dimeric complexes with members of the Dlx homeodomain protein family, preventing DNA binding which then leads to inhibition of their transcriptional activities (Zhang *et al.*, 1997). The crystal structure of the Msx-1–DNA complex should illuminate the structural details of the function of Msx-1 in transcription.

There is a high degree of evolutionary conservation among homeodomains in eukaryotes (Gehring *et al.*, 1994). The conservation present in the homeodomain helped in obtaining a molecular-replacement solution. Several structures of homeodomain–DNA complexes have been solved and, while the majority of experiments indicate that most homeodomains should bind as monomers, most available crystal structures have two protein molecules bound to a single DNA sequence (Kissinger *et al.*, 1990; Hirsch & Aggarwal, 1995), with the exception being the antennapedia homeodomain–DNA complex (Fraenkel & Pabo, 1998). The structure of Msx-1–DNA appears to be a monomer.

2. DNA purification and crystallization

Oligonucleotides were obtained from the Keck Oligonucleotide Synthesis Facility at Yale

Table 1
Data-collection statistics.

Wavelength (Å)	1.54
Resolution range (Å)	40.0–2.15
Reflections recorded	11219
Completeness (last shell) (%)	98.9 (99.8)
$R_{\rm merge}$ (I)† (last shell) (%)	6.2 (29.8)

† $R_{\rm merge} = \sum_I |I_I - \langle I\rangle| / \sum\langle I\rangle$, where I_I is an individual intensity measurement and $\langle I\rangle$ is the average intensity for this reflection, with summation over all data.

University and were purified on an anion-exchange column (Source Q, Pharmacia) on a Perkin–Elmer HPLC with UV detection at 260 nm. 1 µmol of DNA was loaded onto a column (20 ml bed volume) with buffer *A* (10 m*M* NaOH and 0.2 *M* NaCl) and eluted with a shallow gradient of buffer *B* (10 m*M* NaOH and 1.0 *M* NaCl). Fractions were neutralized with 1 *M* Tris at pH 7.5. For concentration, the sample was diluted with four volumes of 10 m*M* Tris and loaded on a 1 ml DEAE cellulose column and the DNA was then eluted with a 1 *M* NaCl elution buffer. The DNA was further concentrated in a Centricon-3 (Amicon) concentrator. The samples were buffer-exchanged into 5 m*M* Tris, 10% glycerol, 50 m*M* KCl and 5 m*M* β-mercaptoethanol. DNA strands were annealed in equimolar amounts. The final DNA concentration was approximately 2 m*M*. The bacterial expression plasmid encoding the homeodomain of Msx-1 and its purification have been discussed in detail elsewhere (Catron *et al.*, 1993). The purity of the protein was determined by SDS–PAGE. The DNA and protein were combined in a 1.2:1 ratio. The final protein concentration was 10 mg ml^{-1}.

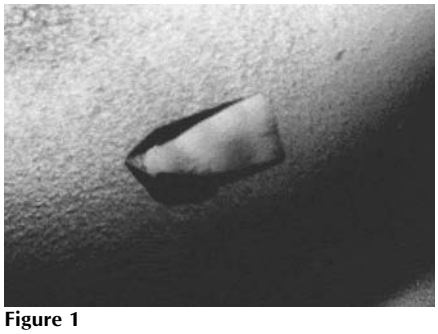

Figure 1
A typical example of an Msx-1 homeodomain–DNA complex crystal. Crystal dimensions were 0.4 × 0.2 × 0.2 mm.

Msx-1–DNA complexes were screened for crystallization in 3 µl hanging drops in a 50-condition sparse-matrix crystal screen (Jancarik & Kim, 1991). Optimization of crystallization yielded diffraction-quality crystals in 12% PEG 4000 and 0.1 *M* sodium acetate pH 4.6. The DNA that yielded the largest crystals was 5′-TGTCACTAATT-GAAGG-3′. It was one of the longer sequences tried (16 base pairs) and contained an overhang T on each end. The crystals grew in four weeks to maximum dimensions 0.4 × 0.2 × 0.2 mm (Fig. 1).

3. Data collection and evaluation

Data was collected from a single crystal at 123 K using an MSC R-AXIS II imaging-plate detector. Cu $K\alpha$ X-rays were produced by a Rigaku RU-200 rotating-anode source operating at 5 kW (50 kV, 100 mA). Data-collection parameters are listed in Table 1. The data was processed with *DENZO* and scaled with *SCALEPACK* (Otwinowski, 1993). The space group was determined to be $P2_12_12_1$, with unit-cell parameters a = 33.66, b = 60.96, c = 83.37 Å, containing one complex in the asymmetric unit (55% solvent). The structure determination is in progress using molecular-replacement methods.

We would like to acknowledge Gezhi Hu and Cory Abate-Shen for providing the protein sample used in crystallization.

References

Catron, K. M., Iler, N. & Abate, C. (1993). *Mol. Cell Biol.* **13**, 2354–2365.
Catron, K. M., Zhang, H., Marshall, S. C., Inostroza, J. A., Wilson, J. M. & Abate, C. (1995). *Mol. Cell Biol.* **15**, 861–871.
Davidson, D. (1995). *Trends Genet.* **11**, 405–411.
Fraenkel, E. & Pabo, C. O. (1998). *Nature Struct. Biol.* **5**, 692–697.
Gehring, W. J., Affolter, M. & Burglin, T. (1994). *Annu. Rev. Biochem.* **63**, 487–526.
Hirsch, J. A. & Aggarwal, A. K. (1995). *EMBO J.* **14**, 6280–6291.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). *Cell*, **63**, 579–590.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M. & Abate-Shen, C. (1997). *Mol. Cell Biol.* **17**, 2920–2932.