STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN: 2053-230X
journals.iucr.org/f

Crystallization and X-ray diffraction analysis of the HMG domain of the chondrogenesis master regulator Sox9 in complex with a ChIP-Seq-identified DNA element

Saravanan Vivekanandan, Balasubramanian Moovarkumudalvan, Julien Lescar and Prasanna R. Kolatkar

Acta Cryst. (2015). F**71**, 1437–1441

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 7 September 2015
Accepted 17 October 2015

Edited by M. L. Pusey, University of Alabama, USA

Keywords: Sox9; transcription factor; HMG domain; FOXP2 promoter; chondrogenesis; sex-determining gene.

Crystallization and X-ray diffraction analysis of the HMG domain of the chondrogenesis master regulator Sox9 in complex with a ChIP-Seq-identified DNA element

Saravanan Vivekanandan,[a,b,c] Balasubramanian Moovarkumudalvan,[a,d] Julien Lescar[b] and Prasanna R. Kolatkar[a,d]*

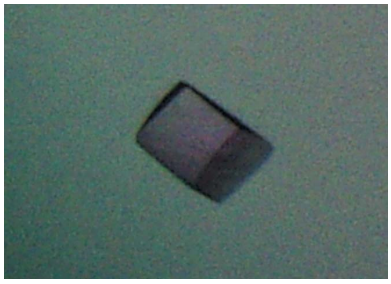
[a]Laboratory for Structural Biochemistry, Genome Institute of Singapore, Genome, 60 Biopolis Street, Singapore 138672, Singapore, [b]School of Biological Science, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore, [c]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore, and [d]Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Qatar Foundation, PO Box 5825, Doha, Qatar. *Correspondence e-mail: pkolatkar@qf.org.qa

Sox9 is a fundamental sex-determining gene and the master regulator of chondrogenesis, and is involved in the development of various vital organs such as testes, kidney, heart and brain, and in skeletal development. Similar to other known Sox transcription factors, Sox9 recognizes and binds DNA with the consensus sequence C(T/A)TTG(T/A)(T/A) through the highly conserved HMG domain. Nonetheless, the molecular basis of the functional specificity of Sox9 in key developmental processes is still unclear. As an initial step towards a mechanistic understanding of Sox9 transcriptional regulation, the current work describes the details of the purification of the mouse Sox9 HMG domain (mSox9HMG), its crystallization in complex with a ChIP-Seq-identified FOXP2 promoter DNA element and the X-ray diffraction data analysis of this complex. The mSox9HMG–FOXP2 promoter DNA complex was crystallized by the hanging-drop vapour-diffusion method using 20% PEG 3350 in 200 m*M* sodium/potassium phosphate with 100 m*M* bis-tris propane at pH 8.5. The crystals diffracted to 2.7 Å resolution and the complex crystallized in the tetragonal space group $P4_12_12$, with unit-cell parameters $a = b = 99.49$, $c = 45.89$ Å. Crystal-packing parameters revealed that asymmetric unit contained one mSox9HMG–FOXP2 promoter DNA complex with an estimated solvent content of 64%.

1. Introduction

Sox [**s**ex-determining region on the Y chromosome (SRY)-b**ox**] transcription factors contain highly conserved SRY-related high-mobility group (HMG) domains of ~80 amino acids that are known to bind and bend DNA. Sox proteins are minor-groove binding (Werner *et al.*, 1995), sequence-specific transcription factors that regulate several key developmental processes. The DNA-binding specificities of the 20 mammalian Sox proteins identified thus far reveal that Sox transcription factors recognize and bind DNA with a C(T/A)TTG(T/A)(T/A) consensus sequence with similar binding preferences (van Houte *et al.*, 1995; Boyer *et al.*, 2005).

Sox proteins are grouped into subfamilies A to J based on the amino-acid sequence similarity of their HMG domains (Kamachi *et al.*, 2000). Sox9, Sox8 and Sox10 belong to group E. Of these, Sox9 is a fundamental sex-determining gene (Clarkson & Harley, 2002) that is involved in the development of various vital organs such as testes, kidney, heart and brain



and in skeletal development. Sox9, partnered by Sox5 and Sox6 from group D, plays a pivotal role as the master regulator of chondrogenesis, regulating multiple stages of cartilage development. Mutations in the *Sox9* gene are known to cause campomelic dysplasia, a skeletal malformation syndrome (Akiyama *et al.*, 2002; Ikeda *et al.*, 2004; Foster *et al.*, 1994). Despite possessing the highly conserved HMG domain involved in degenerate DNA-binding sites, Sox proteins regulate functionally discrete developmental processes. Sox proteins are believed to achieve functional specificity either through structural rearrangement of the HMG-domain arms or by inducing specific kinks in the DNA. Specificity of Sox proteins might also be achieved *via* bending DNA to distinctive degrees, which might subsequently lead to the recruitment of Sox protein-specific cofactors. Interestingly, a comparison of the three published DNA-bound crystal structures of mSox2HMG (Reményi *et al.*, 2003), mSox17HMG (Palasingam *et al.*, 2009) and mSox4HMG (Jauch *et al.*, 2012), belonging to Sox subgroups B, F and C, respectively, reveals that these transcription factors bend DNA to similar extents (∼65°) with a comparable helical bend axis, and preserve the characteristic L-shaped fold of their helices with little structural rearrangement. The few available Sox HMG-domain structures limit our understanding of the functional specificity of the Sox transcription factors. Therefore, high-resolution structure determination of the various Sox-subgroup HMG domains would provide a comprehensive insight into the mechanism of Sox transcriptional regulation.

To this end, we have attempted to determine the DNA-bound HMG-domain structure of mouse Sox9 (mSox9HMG), the master regulator of chondrogenesis. Previous Sox HMG-domain crystal structures have employed DNA elements derived from known Sox enhancer elements such as LAMA1 and FGF4. Alternatively, in the current work, in order to better model the precise *in vivo* functional binding sites of Sox9, we employed a Sox9-specific DNA element derived from the *FOXP2* gene promoter, identified through immunoprecipitation coupled with ultrahigh-throughput DNA sequencing (ChIP-Seq; Professor Thomas Lufkin, Genome Institute Singapore, private communication). Here, we present the protein expression, purification, crystallization and preliminary X-ray diffraction data of mSox9HMG bound to a FOXP2-derived 16-mer promoter DNA element.

2. Materials and methods

2.1. Cloning and expression

The 80-amino-acid HMG domain of mouse Sox9, spanning residues 103–183 of the full-length protein, was PCR-amplified from the IMAGE 5354229 cDNA clone using the gene-specific primers 5′-CACCCCACACGTCAAGCGACC-3′ and 5′-TTACACCGACTTCCTCCGCCG-3′. The amplified PCR product was cloned into pENTR/TEV/D-TOPO by a directional TOPO cloning (Invitrogen) vector to generate entry clones, which were further verified by colony PCR and DNA sequencing. The *mSox9HMG* gene in the entry clone was introduced into the Gateway destination vector pETG20A by performing a Gateway LR reaction, yielding the pETG20A-mSox9HMG expression plasmid, and the presence of the gene was validated by PCR using gene-specific primers. The expression plasmid pETG20A-mSox9HMG thus obtained was transformed into *Escherichia coli* BL21 (DE3) cells (Invitrogen), which were grown in Luria–Bertani (LB) broth containing 100 µg ml^{-1} ampicillin and 0.2% glucose at 310 K to an OD_{600} of 0.7. The temperature was lowered to 303 K and protein expression was induced by the addition of 0.3 m*M* IPTG. The cells were harvested after 4 h by centrifugation using an SLA-3000 fixed-angle rotor at 10 000 rev min^{-1} for 10 min and stored at 193 K.

2.2. Protein purification

The harvested cells were thawed, resuspended in buffer *A* (20 m*M* HEPES pH 7.0, 1 m*M* EDTA, 100 m*M* NaCl) and disrupted by ultrasonication for 20 min on ice. The lysate was clarified by centrifugation using an SS-34 fixed-angle rotor at 18 000 rev min^{-1} for 30 min and passed through a 0.22 µm filter. The supernatant was incubated with Ni–NTA agarose beads (Qiagen) pre-equilibrated with buffer *A*, and the Thx-His_6-mSox9HMG was purified using buffer *A* with an imidazole gradient (25–300 m*M*). The purified Thx-His_6-mSox9HMG was subjected to cleavage by TEV protease and purified by ion-exchange chromatography (RESOURCE S, volume 6 ml; GE Healthcare) using a gradient of buffer *B* (20 m*M* HEPES pH 7.0, 1 m*M* EDTA, 1 *M* NaCl). The mSox9HMG protein was further purified to homogeneity by size-exclusion chromatography using a Superdex 75 column. The purity of the appropriate protein peak fractions was assessed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and SDS–PAGE. Pooled fractions were concentrated to 5–10 mg ml^{-1} as estimated by standard protein absorbance (A_{280}) using a Thermo Scientific NanoDrop ND-1000 spectrophotometer.

2.3. Crystallization of mSox9 with FOXP2 promoter DNA

Single-stranded DNA oligonucleotides with varying overhangs purified *via* polyacrylamide gel electrophoresis were commercially obtained at a concentration of 1000 µ*M* (Proligo, Sigma–Aldrich). The complementary oligonucleotides were mixed in equimolar concentrations and annealed by heating to 95°C and gradually cooling to ambient temperature. The purified mSox9HMG and the double-stranded DNA were mixed in a 1:1.2 molar ratio and further incubated on ice for 2 h. The mSox9HMG–FOXP2 promoter DNA complex thus formed was subjected to crystallization trials at a protein concentration of ∼320 µ*M*. Optimal crystal-growth conditions were screened with commercial crystal screening kits from Hampton Research and Qiagen using a liquid-dispensing robot (Innovadyne). Crystallization trials were carried out using the sitting-drop vapour-diffusion method by combining equal volumes of protein solution and precipitating buffer. Optimization of the conditions was carried out by varying the

lengths and the overhangs of the DNA and by using different ratios of concentrations of the protein and precipitants.

2.4. X-ray data collection and processing

Crystals were flash-cooled in liquid nitrogen and a 2.7 Å resolution native data set was collected on beamline X29A at Brookhaven National Synchrotron Light Source (NSLS), New York, USA at a wavelength of 1.0750 Å. A total of 360 images were collected, each with an oscillation angle of 1°. Diffraction intensities were collected using an ADSC Q315 CCD detector and were processed and scaled using the *HKL*-2000 suite (Otwinowski & Minor, 1997).

2.5. Electrophoretic mobility-shift assay (EMSA)

EMSA experiments were performed by incubating 5′ Cy5-labelled (Proligo, Sigma–Aldrich) 16-mer dsDNA FOXP2 promoter elements with mSox9HMG in a binding buffer consisting of 20 m*M* Tris–HCl pH 8.0, 0.1 mg ml^{-1} bovine serum albumin, 50 µ*M* $ZnCl_2$, 100 m*M* KCl, 10% glycerol, 0.1% NP-40, 2 m*M* β-mercaptoethanol. The protein–DNA complex was formed by incubating 0.1, 0.5, 0.75, 1, 1.5, 2, 5, 10 and 12.5 n*M* protein with 1 n*M* probe for 1 h at 4°C in the dark in a 10 µl reaction volume. Samples were loaded onto 12% 1× TG native polyacrylamide gels and were electrophoresed in 1× TG (25 m*M* Tris pH 8.3, 192 m*M* glycine) at 200 V for 30 min at 4°C. Bands were detected using a Typhoon 9140 phosphorimager (GE Healthcare).

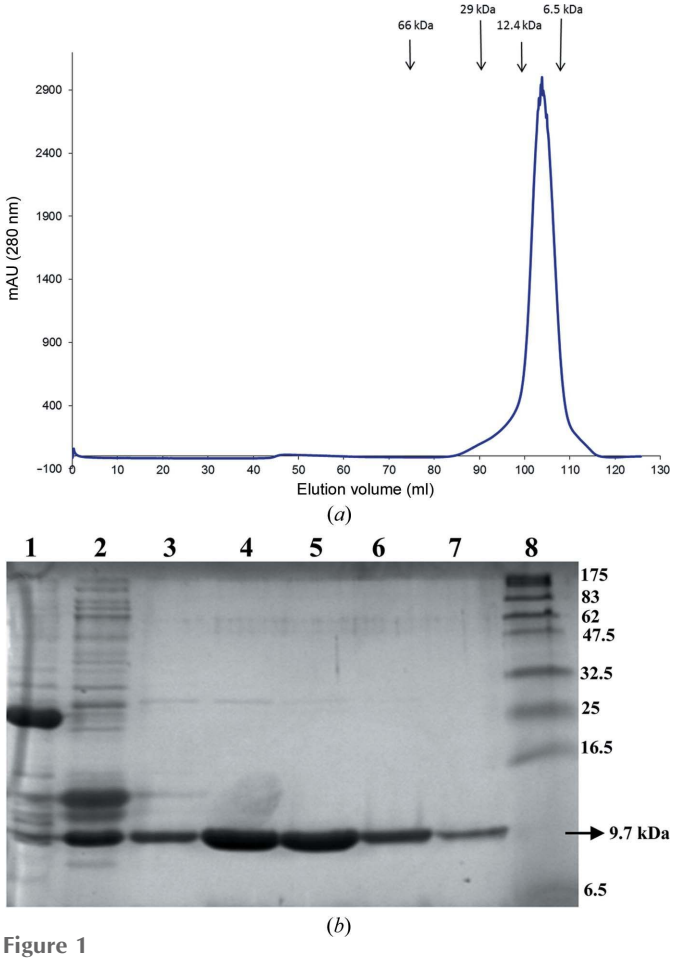


Figure 1
(*a*) mSox9HMG elution profile from a Superdex 75 column, showing a single symmetric peak corresponding to a molecular weight of ∼9.7 kDa compared with molecular-weight standards [bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome *c* (12.4 kDa) and aprotinin (6.5 kDa)]. (*b*) 12% SDS–PAGE analysis. Lane 1, uncleaved mSox9HMG protein; lane 2, cleaved fraction with both thioredoxin tag and the protein of interest; lanes 3–7, elution fractions from Superdex 75 consistent with a molecular weight of ∼9.7 kDa; lane 8, molecular-weight markers (labelled in kDa).

3. Results and discussion

3.1. Protein preparation and protein–DNA complex formation

mSox9HMG protein was overexpressed in a bacterial expression system at an optimal temperature of 30°C after induction with 0.3 m*M* IPTG and was purified in a soluble form to homogeneity, with typical yields of ∼3.5 mg per litre. The pure mSox9HMG protein eluted as a monomer with an apparent molecular mass of 9.7 kDa from a gel-filtration column (Fig. 1*a*), and an SDS–PAGE analysis showed >98% purity (Fig. 1*b*).

The *in vivo* data generated by immunoprecipitation coupled with ultrahigh-throughput DNA sequencing (ChIP-Seq) of chondrogenic limb and tail tissues of germline-transmitting (GLT) chimeras from Sox9^{+}-(EGFP) mouse, utilized to identify and validate the regulatory motifs in cartilage-specific genes, yielded a novel Sox9 consensus binding sequence of 5′-AGAACAAAG-3′, corresponding to the *FOXP2* gene promoter sequence. EMSA using Cy5-labelled dsDNA harbouring the FOXP2 motif sequence and mSox9HMG revealed a very high binding affinity, with a dissociation constant (K_d) of ∼1.4 n*M* (Fig. 2). In contrast to Sox9 binding motifs reported previously based on computational and *in vitro* approaches, the unpublished ChIP-Seq *in vivo* results used here identify a precise and reliable transcription-factor binding site and hence justify the use of the FOXP2 promoter DNA element for crystallization with mSox9HMG.

3.2. Crystallization

Crystallization trials for the homogenously purified mSox9HMG–FOXP2 promoter DNA complex were set up using the sitting-drop vapour-diffusion method. Initial

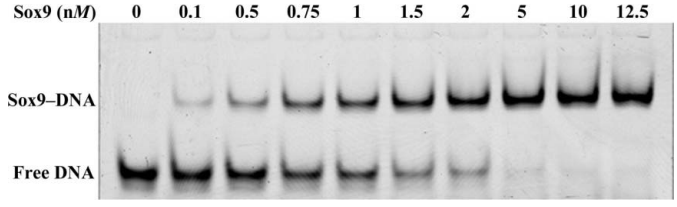
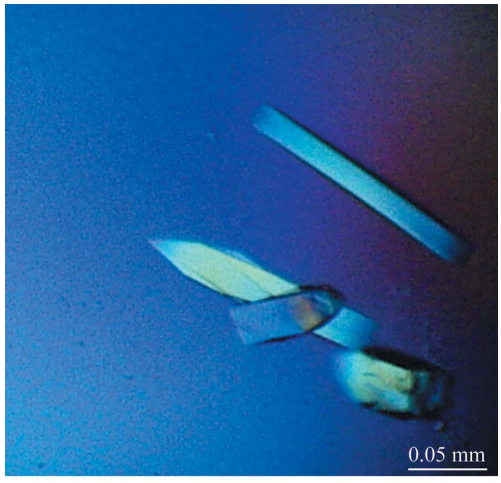
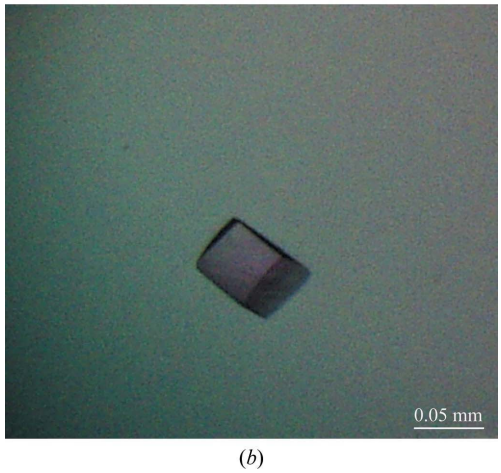


Figure 2
An EMSA experiment with titration of the mSox9HMG protein with FOXP2 promoter DNA. 1 n*M* Cy5-labelled FOXP2 dsDNA incubated with 0, 0.1, 0.5, 0.75, 1, 1.5, 2, 5, 10 and 12.5 n*M* mSox9HMG protein and subjected to an EMSA experiment revealed a very high binding affinity, with a dissociation constant (K_d) of ∼1.4 n*M*.

co-crystallization of mSox9HMG with a blunt-ended 17-mer derived from the *FOXP2* gene promoter sequence yielded crystals that were either fragile or of poor diffraction quality. The length of the DNA element and the number of unpaired base pairs in the flanking region are two parameters that are routinely varied to obtain good-quality crystals of protein–DNA complexes. Consequently, various FOXP2 oligonucleotides ranging from 16-mers to 17-mers and with AT/CG/GC/GG/CC overhangs were utilized for protein–DNA complex formation and the effect on crystal formation was analysed (Table 1). Of all the variants, crystals formed using 17-mer blunt-ended DNA diffracted to only 9 Å resolution, whereas those formed using 17-mer CG, 16-mer GC and 16-mer CC overhangs diffracted to 6 Å resolution. The use of a 16-mer AT overhang did not yield any crystals. Only a 16-mer GG overhang yielded better crystals, in the presence of 16% PEG 3350 in 2% Tacsimate pH 5.0 with 100 m*M* trisodium citrate pH 5.6, which diffracted to 3 Å resolution (Fig. 3*a*). However, data processing was hindered owing to high mosaicity. Finally, good diffraction-quality crystals of a complex with a 16-mer FOXP2 DNA (5′-AGGAGAACAAAGCCTG-3′) containing a GG overhang were obtained at 18°C using a 1:2 ratio of protein–DNA complex at a concentration of 257 µ*M* to a reservoir solution consisting of 20% PEG 3350 in 200 m*M* sodium/potassium phosphate with 100 m*M* bis-tris propane pH 8.5 (Fig. 3*b*). The crystals were harvested after 25 d, flash-cooled and stored in liquid nitrogen. In addition, the crystals were washed with mother liquor, dissolved in buffer and subjected to SDS–PAGE; the agarose gel analysis revealed the presence of both protein and DNA.



(*a*)



(*b*)

Figure 3
Crystals of mSox9HMG–FOXP2 promoter DNA with a GG overhang (5′-AGGAGAACAAAGCCTG-3′) grown in buffers consisting of (*a*) 16% PEG 3350 in 2% Tacsimate pH 5.0 with 100 m*M* trisodium citrate pH 5.6, which diffracted to 3 Å resolution, and (*b*) 20% PEG 3350 in 200 m*M* sodium/potassium phosphate with 100 m*M* bis-tris propane pH 8.5, which diffracted to 2.7 Å resolution.

Table 1
DNA elements derived from the FOXP2 promoter region used in crystallization trials.

DNA element used	Sequence	Crystals	Resolution (Å)
Foxp2_17_Blunt	5′-CAGGAGAACAAAGCCTG-3′ 3′-GTCCTCTTGTTTCGGAC-5′	Y	∼9
Foxp2_17_CG	5′-CAGGAGAACAAAGCCTG-3′ 3′-TCCTCTTGTTTCGGACG-5′	Y	∼6
Foxp2_16_GC	5′-GAGGAGAACAAAGCCT-3′ 3′-TCCTCTTGTTTCGGAC-5′	Y	∼6
Foxp2_16_CC	5′-CAGGAGAACAAAGCCT-3′ 3′-TCCTCTTGTTTCGGAC-5′	Y	∼6
Foxp2_16_AT	5′-AGGAGAACAAAGCCTG-3′ 3′-CCTCTTGTTTCGGACT-5′	N	NA
Foxp2_16_GG	5′-AGGAGAACAAAGCCTG-3′ 3′-GTCCTCTTGTTTCGGA-5′	Y	2.7

Table 2
Data-collection and processing statistics for the mSox9HMG–FOXP2 promoter DNA complex.

Values in parentheses are for the highest resolution bin.

Source	X29A, NSLS
Wavelength (Å)	1.0750
Space group	$P4_12_12$
Unit-cell parameters (Å, °)	$a = b = 99.49$, $c = 45.89$, $\alpha = \beta = \gamma = 90$
Resolution range (Å)	50.0–2.70 (2.80–2.70)
Total No. of reflections	203327
Unique reflections	6721
$R_{\rm merge}$† (%)	13.3 (65.4)
Average multiplicity	30.3 (28.2)
Completeness (%)	99.9 (100.0)
Average $I/\sigma(I)$	19.0 (6.9)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the measured intensity of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity.

3.3. Data collection and processing

The optimized crystals of the mSox9HMG–FOXP2 promoter DNA complex diffracted to 2.7 Å resolution and the crystals belonged to the tetragonal space group $P4_12_12$ or its enantiomorph $P4_32_12$, with unit-cell parameters $a = b = 99.49$, $c = 45.89$ Å. Crystal-packing parameters revealed the presence of one mSox9HMG–FOXP2 promoter DNA complex per asymmetric unit, with a Matthews coefficient of 2.8 Å^3 Da^{-1} and a solvent content of 64% (Matthews, 1968). The data-collection and data-processing statistics are presented in Table 2. Initial phase determination was attempted by molecular replacement with the mSox17HMG–LAMA1 DNA complex (PDB entry 3f27; Palasingam *et al.*, 2009) as a starting

model using *Phaser* (McCoy *et al.*, 2005) as implemented in the *CCP*4 suite (Winn *et al.*, 2011). Further model building and refinement of the structure is in progress.

Acknowledgements

This work was supported by the Agency for Science, Technology and Research (A*STAR) in Singapore. We are grateful to Professor Thomas Lufkin (GIS, Singapore) for the ChIP-Seq data. We thank Dr Robert Robinson (IMCB, Singapore) for access to the in-house robotics facility and diffraction data collection. We also thank Dr Howard Robinson at beamline X29A of the National Synchroton Light Source, Brookhaven National Laboratory for providing assistance in data collection. SV is also grateful to NTU for providing a Nanyang Graduate scholarship.

References

Akiyama, H., Chaboissier, M.-C., Martin, J. F., Schedl, A. & de Crombrugghe, B. (2002). *Genes Dev.* **16**, 2813–2828.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A. (2005). *Cell*, **122**, 947–956.

Clarkson, M. J. & Harley, V. R. (2002). *Trends Endocrinol. Metab.* **13**, 106–111.

Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., Kwok, C., Weller, P. A., Stevanović, M., Weissenbach, J., Mansour, S., Young, I. D., Goodfellow, P. N., Brook, J. D. & Schafer, A. J. (1994). *Nature (London)*, **372**, 525–530.

Houte, L. P. A. van, Chuprina, V. P., van der Wetering, M., Boelens, R., Kaptein, R. & Clevers, H. (1995). *J. Biol. Chem.* **270**, 30516–30524.

Ikeda, T., Kamekura, S., Mabuchi, A., Kou, I., Seki, S., Takato, T., Nakamura, K., Kawaguchi, H., Ikegawa, S. & Chung, U. (2004). *Arthritis Rheum.* **50**, 3561–3573.

Jauch, R., Ng, C. K. L., Narasimhan, K. & Kolatkar, P. R. (2012). *Biochem. J.* **443**, 39–47.

Kamachi, Y., Uchikawa, M. & Kondoh, H. (2000). *Trends Genet.* **16**, 182–187.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Palasingam, P., Jauch, R., Ng, C. K. L. & Kolatkar, P. R. (2009). *J. Mol. Biol.* **388**, 619–630.

Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R. & Wilmanns, M. (2003). *Genes Dev.* **17**, 2048–2059.

Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995). *Cell*, **81**, 705–714.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.