Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

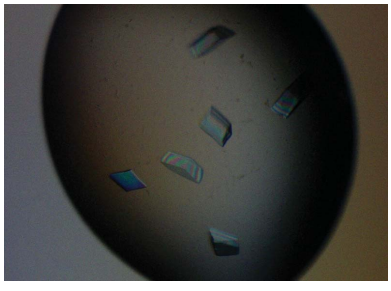
Crystallization and preliminary X-ray diffraction analysis of the Pax9 paired domain bound to a DC5 enhancer DNA element

Kamesh Narasimhan,[a,b]* Antonia Hilbig,[b,c] Barath Udayasuryan,[b,d] Sriram Jayabal,[b,e] Prasanna R. Kolatkar[b,f] and Ralf Jauch[b,g]*

[a]Donnelly Centre for Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada, [b]Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore, [c]Institute für Biochemie, Freie Universität Berlin, Thielalle 63, 14195 Berlin, Germany, [d]Department of Biomedical Engineering, University of Michigan, 1107 Carl A. Gerstacker Building, 2200 Bonisteel Boulevard, Ann Arbor, MI 48109-2099, USA, [e]Integrated program in Neuroscience, McGill University, 3649 Sir-William-Osler, Montreal, Quebec H3G 0B1, Canada, [f]Qatar Biomedical Research Institute, Qatar Foundation, PO Box 5825, Doha, Qatar, and [g]Genome Regulation Laboratory, Guangzhou Insitutes of Biomedicine and Health, 190 Kai Yuan Avenue, Science Park, Guangzhou 510530, People's Republic of China

Correspondence e-mail: kamesh.narasimhan@utoronto.ca, ralf@gibh.ac.cn

Received 6 May 2014
Accepted 29 July 2014



Pax genes belong to a family of metazoan transcription factors that are known to play a critical role in eye, ear, kidney and neural development. The mammalian Pax family of transcription factors is characterized by a ~128-amino-acid DNA-binding paired domain that makes sequence-specific contacts with DNA. The diversity in Pax gene activities emerges from complex modes of interaction with enhancer regions and heterodimerization with multiple interaction partners. Based on *in vitro* optimal binding-site selection studies and enhancer identification assays, it has been suggested that Pax proteins may recognize and bind their target DNA elements with different binding modes/topologies, however this hypothesis has not yet been structurally explored. One of the most extensively studied DNA target elements of the Pax6 paired domain is the eye-lens specific DC5 (δ-crystallin) enhancer element. In order to shed light on Pax6–DC5 DNA interactions, the related paired-domain prototype Pax9 was crystallized with the minimal δ-crystallin DC5 enhancer element and preliminary X-ray diffraction analysis was attempted. A 3.0 Å resolution native data set was collected at the National Synchrotron Light Source (NSLS), Brookhaven from crystals grown in a solution consisting of 10%(*w*/*v*) PEG 20K, 20%(*v*/*v*) PEG 550 MME, 0.03 *M* $NaNO_3$, 0.03 *M* Na_2HPO_4, 0.03 *M* NH_2SO_4, 0.1 *M* MES/imidazole pH 6.5. The data set was indexed and merged in space group $C222_1$, with unit-cell parameters $a = 75.74$, $b = 165.59$, $c = 70.14$ Å, $\alpha = \beta = \gamma = 90^\circ$. The solvent content in the unit cell is consistent with the presence of one Pax9 paired domain bound to duplex DNA in the asymmetric unit.

1. Introduction

The Pax family of transcription factors (TFs) is characterized by the presence of an ~128-amino-acid paired domain (PD) that is known to bind DNA in a sequence-specific manner (Noll, 1993). The paired domain has a bipartite arrangement composed of an N-terminal subdomain (NTD) and a C-terminal subdomain (CTD) joined by a linker region (Xu *et al.*, 1999; Balczarek *et al.*, 1997). The Pax proteins mediate a panoply of functions in haematopoiesis, neurogenesis and eye development by engaging in alternative DNA-recognition modes and participating in complex heterodimerization partnerships with the Sox, Ets and Homeodomain families (Kamachi *et al.*, 2001; Garvie *et al.*, 2001; Li-Kroeger *et al.*, 2012). Juxtaposition of the optimal Pax6 DNA sequence from *in vitro* DNA-binding site-selection experiments with a number of paired-domain responsive DNA sequences reveals that there are significant deviations in binding-sequence preferences within the Pax genes (Kamachi *et al.*, 2001; Inoue *et al.*, 2007; Epstein *et al.*, 1994; Pellizzari *et al.*, 1999).

The crystal structure of the Pax6 PD with the optimal Pax6 binding site reveals that both the N-terminal and C-terminal domains consist of three compact α-helices, with an additional β-hairpin and β-turn in the N-terminal domain (Xu *et al.*, 1999). The third helix of the N-terminal and C-terminal domains docks against the major groove, while the linker region makes extensive DNA contacts with the minor groove. The N-terminal β-hairpin makes sugar-phosphate backbone contacts, while the β-turn makes specific base contacts with the minor groove (Xu *et al.*, 1999). It can be hypothesized that the Pax proteins recognize alternate DNA sequences through either the use of different combinations of DNA-binding residues while maintaining a

similar protein conformation or by binding to the sequences with a different topology in order to accommodate sequences. Specifically, we wanted to understand the differential DNA-recognition mechanism of the Pax6 PD in binding to the eye-lens specific DC5 enhancer element and thus attempted to solve the structure of this complex. The Pax9 PD exhibits an overall sequence identity of ∼69% to the Pax6 PD, with a high degree of conservation at the level of DNA-binding residues, and was thus used as a proxy for the biologically relevant Pax6, as attempts to crystallize Pax6 with the DC5 DNA resulted in fragile and poor-quality crystals (Fig. 1*a*).

Table 1
Primers used for cloning and the final amino-acid sequence of the Pax9 PD used for crystallization.

*att*B sites are underlined, the TEV cleavage site is italicized and Pax9 sequence-specific sites are highlighted in bold.

Forward primer	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTC*GAAAACCTGT-ATTTTCAGGGC*GCCTTCGGGGAGGTGAACCAGCTG-3′
Reverse primer	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTAGTTGCCGAT-**CTTGTTGCGCAGAATAC**-3′
Amino-acid sequence of the construct	GAFGEVNQLGGVFVNGRPLPNAIRLRIVELAQLGIRPCDISRQL-RVSHGCVSKILARYNETGSILPGAIGGSKPRVTTPTVVKHI-RTYKQRDPGIFAWEIRDRLLADGVCDKYNVPSVSSISRILR-NKIGN

2. Materials and methods

2.1. Cloning, expression and purification of the Pax9 PD

The Pax9 PD spanning amino acids 4–133 of the full-length mouse Pax9 protein was PCR-amplified from the Pax9 cDNA (IMAGE 3707718) using primers containing *att*B and TEV protease cleavage sites (Table 1). The PCR product was first cloned into pDONR221 (entry vector) and later transferred into pETG40A (destination vector) encoding an MBP fusion tag using Gateway cloning (Invitrogen). The pETG40A Pax9 PD expression plasmid was transformed into competent *Escherichia coli* BL21 (DE3) cells and the transformed cells were grown overnight at 37°C in LB medium containing 100 µg ml^{-1} ampicillin. A 10%(*v*/*v*) dilution of the overnight culture was used to inoculate a 5 l LB broth culture at 37°C. When the optical density (OD_{600}) of the culture reached 0.6, the cells were induced with 0.5 m*M* IPTG and then grown continuously at 18°C overnight. The cells were then pelleted and resuspended in lysis buffer (40 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5 m*M* *β*-mercaptoethanol, 2 m*M* EDTA) and sonicated on ice for 15 min. Fusion proteins were extracted from the cell lysate using an amylose column equilibrated with lysis buffer and eluted with the same buffer supplemented with 10 m*M* maltose. The fusion MBP tag was cleaved using TEV protease at 4°C overnight [substrate:TEV ratio of

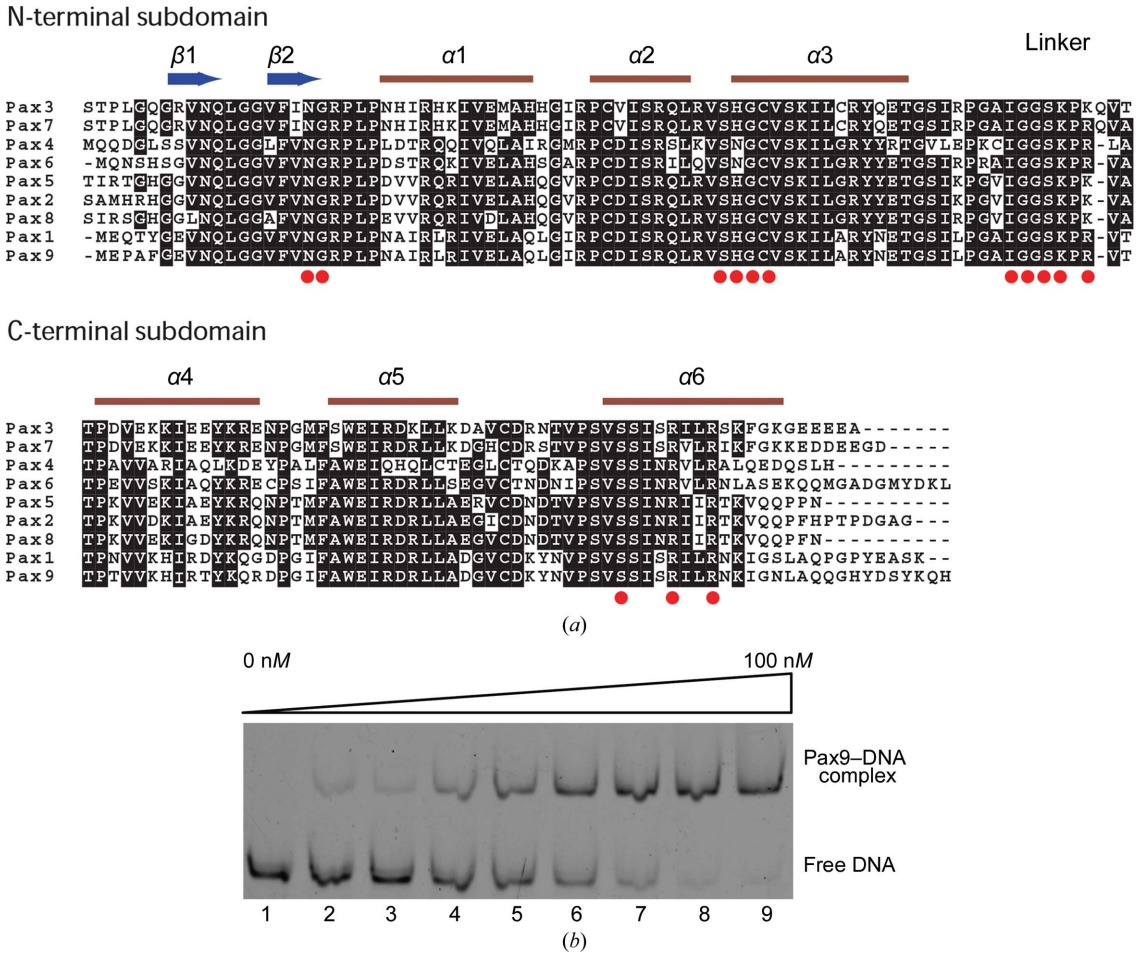


Figure 1
(*a*) Multiple sequence alignment of the Pax9 PD with other mouse Pax family members was carried out using *ClustalW*. Key amino-acid residues of the paired domain (PD) involved in major-groove and minor-groove DNA interactions are highlighted using red spheres. (*b*) EMSA was carried out with 1 n*M* 5′-FAM-labelled DC5 DNA incubated with varying concentrations of Pax9 PD. The lowest band in lane 1 corresponds to 1 n*M* of the free DNA without any protein. Protein concentrations increase in a twofold manner from left to right to a final concentration of 100 n*M* (lanes 2–9)

Table 2
Final crystallization conditions.

Method	Hanging drop
Plate type	EasyXtal 15-Well Tool
Temperature (°C)	18
Protein concentration (mg ml^{-1})	8
Buffer composition of protein and DNA solution	40 m*M* HEPES pH 7.5, 100 m*M* NaCl, 5 m*M* β-mercaptoethanol
Volume and ratio of drop	1:1 (1 µl:1 µl)
Volume of reservoir (µl)	500
DNA element† (DC5_T/A)	5′-*TTT*GTTGCTCACCTACCATGGACAAT-3′ 3′-AAC*AA*CGAGTGGATGGTACCTGTT*AA*-5′

† The DC5 core recognition site is underlined. Overhangs in the DNA element are italicized.

100:1(*w*:*w*)]. Further purification with a heparin column (GE) was performed using a linear gradient ranging from 100 m*M* to 1.0 *M* NaCl to remove the MBP and TEV fusion tags. Pax9 was eluted and subjected to a final purification step using HiPrep S-75 gel-filtration chromatography in gel-filtration buffer (40 m*M* HEPES pH 7.5, 100 m*M* NaCl, 5 m*M* β-mercaptoethanol). Pax9 fractions were concentrated to 100 µ*M* by centrifugation and the purity of the collected fraction was verified on an SDS–PAGE gel.

2.2. Electrophoretic mobility shift assay (EMSA) of the Pax9 PD with the DC5 element

The activity of the recombinantly purified Pax9 PD in binding to the DC5 element was established by EMSA prior to setting up large-scale crystallization screens. Pax9 PD was titrated at varying concentrations from 0.78 to 100 n*M* (in a twofold dilution series) with 1 n*M* of a fluorescein (FAM)-labelled DC5 DNA element and incubated for 1 h. The samples were then loaded onto a 10% native polyacrylamide gel and electrophoresis was carried out at 200 V for 20 min at 4°C using a previously established EMSA protocol (Baburajendran *et al.*, 2011). The gel was subsequently imaged using a Typhoon phosphorimaging scanner. The forward-strand sequence of the duplex DC5 DNA used was 5′-(FAM)-TTTG**TTGCTCACCTACCATGGACAA**T-3′ (the core Pax9 PD binding site is shown in bold). The Pax9 PD bound to the DC5 DNA element, as confirmed by the retarded migration of the protein–DNA complex in EMSA (Fig. 1*b*).

2.3. Preparation of duplex DC5 DNA for crystallization

Polyacrylamide gel electrophoresis (PAGE)-purified, deprotected single-stranded DNA oligonucleotides were purchased at 1 m*M* concentration predissolved in molecular-grade water (Proligo, Sigma–Aldrich). Equimolar amounts of complementary DNAs were mixed and concentrated using a Centricon filter unit (3 kDa molecular-weight cutoff) into a buffer composed of 40 m*M* HEPES pH 7.5, 5 m*M* β-mercaptoethanol, 100 m*M* NaCl to a final oligonucleotide concentration of 500 µ*M* as measured using a NanoDrop spectrophotometer. The oligonucleotide mixture was then heated to 95°C for 5 min and slowly cooled at a rate of 0.5°C min^{-1} to 25°C in a thermocycler and stored at −30°C.

2.4. Crystallization

Pax9 PD was mixed with the the DC5_T/A DNA element containing the core δ-crystallin binding site in a molar ratio of 1:1.2 and incubated for 1 h on ice. The Pax9 PD–DNA complex was concentrated using a Centricon filter unit (3 kDa molecular-weight cutoff) to a final protein concentration of 8 mg ml^{-1} as estimated using the Bradford reagent. Preliminary high-throughput screening was carried out with an Innovadyne robot in 96-well sitting-drop plates using a number of chemical screens from the following manufacturers: PEG/Ion and Nucleix from Hampton Research, The PEGs, PEGs II, $AmSO_4$, PACT and JCSG+ Suites from Qiagen, Morpheus from Molecular Dimensions and JB-NucPro from Jena Biosciences. The high-throughput sitting-drop screens were carried out by combining 200 nl protein–DNA complex with 200 nl reservoir solution over a 50 µl reservoir, with the plates being stored at 18°C. Among the different conditions that were screened, it was observed that consistently better quality crystals were obtained in many conditions from Morpheus.

Several optimizations were then performed by the hanging-drop method around the initial Morpheus hits by varying the DNA element length and overhangs, the temperature (25, 18 and 4°C), the pH and the salt and PEG constituents. The crystals were harvested within 2 d, soaked in 15% glycerol as a cryoprotectant for less than 5 s and flash-cooled in liquid nitrogen for data collection. The final diffraction-quality crystal from which the complete data set was collected was obtained by the hanging-drop method in the presence of 10%(*w*/*v*) PEG 20 K, 20%(*v*/*v*) PEG 550 MME, 0.03 *M* $NaNO_3$, 0.03 *M* Na_2HPO_4, 0.03 *M* NH_2SO_4, 0.1 *M* MES–imidazole pH 6.5 at

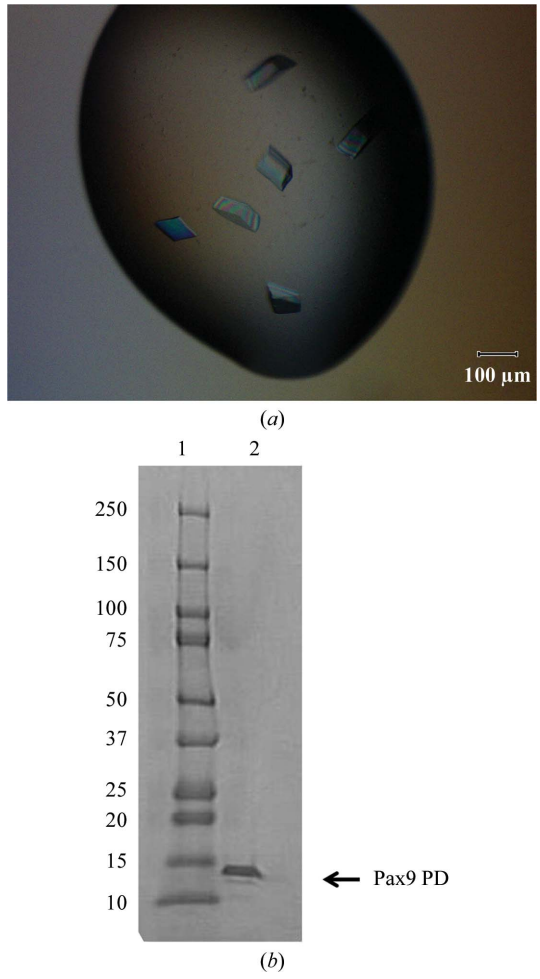


Figure 2
(*a*) Diamond-shaped rhombic crystals of the Pax9 PD–DC5_T/A complex grown at 18°C that gave the best-quality diffraction to 3.0 Å resolution. (*b*) Several crystals of the Pax9 PD–DC5 DNA complex were washed in reservoir solution, crushed, dissolved and run on 12% SDS–PAGE stained with SimplyBlue Stain (Invitrogen), showing a band corresponding to the expected Pax9 PD protein at 14.28 kDa in lane 2. Lane 1 was loaded with molecular-weight marker (labelled in kDa).

18°C with the DC5_T/A DNA element (Table 2, Fig. 2*a*). The presence of the protein in the crystal was confirmed by analyzing dissolved crystals using SDS–PAGE as described in Ng *et al.* (2008) (Fig. 2*b*).

2.5. Data collection and processing

The initial X-ray diffraction tests were performed in-house using a PLATINUM[135] CCD detector with focused Cu $K\alpha$ X-rays from an X8 PROTEUM rotating-anode generator (Bruker AXS) controlled by the *PROTEUM2* software (Sheldrick, 2008). The final 3.0 Å resolution native data set was collected on the X29A beamline at the National Synchrotron Light Source (NSLS), Brookhaven. A total of 360 images were collected with a crystal-to-detector distance of 300 mm, an oscillation angle of 1° and an exposure time of 0.4 s per image. The data set was integrated, merged and scaled using the *HKL*-2000 software (Otwinowski & Minor, 1997). The data-collection and processing statistics for the Pax9 PD–DC5 complex are provided in Table 3.

Table 3
Data-collection and processing statistics for the Pax9 PD–DC5 complex.

Values in parentheses are for the outer shell.

Diffraction source	X29A, NSLS
Wavelength (Å)	1.075
Temperature (°C)	−168
Detector	ADSC Quantum 315r
Space group	$C222_1$
Unit-cell parameters (Å, °)	a = 75.74, b = 165.59, c = 70.14, $\alpha = \beta = \gamma$ = 90
Resolution range (Å)	50–3.0 (3.11–3.00)
Total No. of reflections	119249 (5321)
No. of unique reflections	8557 (606)
Completeness (%)	94.93 (73.45)
Multiplicity	13.7 (8.6)
$\langle I/\sigma(I)\rangle$	12.99 (2.94)
$R_{\rm meas}$† (%)	5.5 (48.1)

† $R_{\rm meas}$ was estimated by multiplying the conventional $R_{\rm merge}$ value by the factor $[N/(N-1)]^{1/2}$, where N is the data multiplicity

3. Results and discussion

It was observed that the Pax9 PD–DC5 crystals invariably tend to exhibit anisotropic diffraction with a resolution range of ∼3–6 Å in a variety of screening conditions. To reduce the anisotropic diffraction, several combinations of optimization methods such as the Additive and Silver Bullets Screens (Hampton Research) and crystal-dehydration methods were undertaken. Crystals grown with additives and Silver Bullets tended to have sharper edges and grew to slightly larger than their counterparts. However, this optimization method failed to significantly reduce the anisotropic diffraction. Crystal dehydration was carried out under hanging-drop conditions by increasing the concentration range of the precipitants [10–15%(*w*/*v*) PEG 20K and 20–30%(*v*/*v*) PEG 550 MME] in the reservoir. Crystals dehydrated for longer than 2 h tended to become fragile, while dehydration for shorter intervals also did not improve the overall diffraction quality. Flash-annealing by diverting the cryostream for 2–3 s was also unsuccessful. Although many crystallization conditions were optimized and the resulting crystals tested for diffraction at the synchrotron, the best crystal still diffracted anisotropically to ∼3.0 Å resolution (Fig. 3).

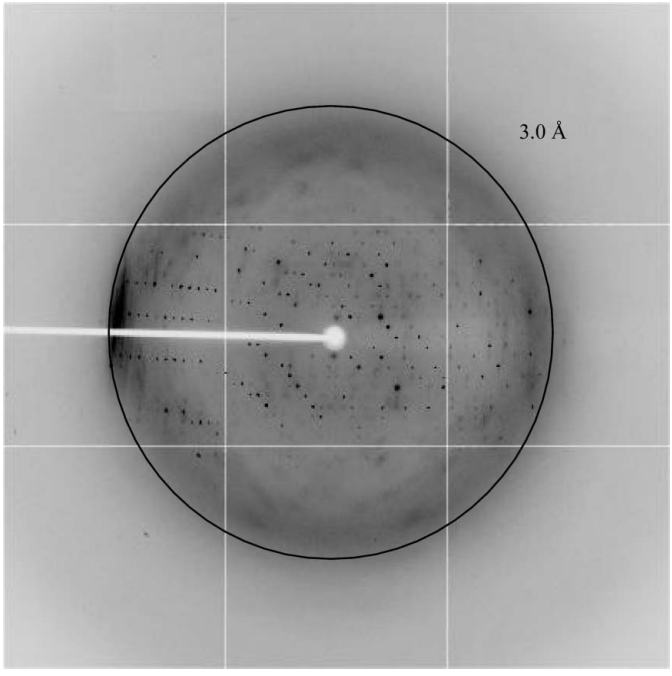


Figure 3
X-ray diffraction pattern of the best-quality Pax9 PD–DC5_T/A crystal recorded at NSLS, Brookhaven at a wavelength of 1.075 Å with a crystal-to-detector distance of 300 mm and an oscillation angle of 1°. The image shows anistropic diffraction with a diffraction limit of ∼3.0 Å resolution.

The Matthews coefficient for the final data set was 3.67 Å^3 Da^{-1}, suggesting a solvent content of 70.06%, assuming that one Pax9 PD domain is bound to the DC5 enhancer DNA (Matthews, 1968). The observation of such a relatively high solvent content could be indirectly deduced from the fragile nature of the crystals as well as from the persistent anisotropic diffraction pattern (Fig. 3). A model derived from the Pax6 structure (PDB entry 6pax; Xu *et al.*, 1999) was unsuccessfully used for molecular-replacement trials. It must be noted that several previous attempts to obtain better diffraction-quality crystals for the homologous Pax8 and Pax6 paired domains on their cognate thyroglobulin promoter and LE9 elements have been unsuccessful as well (Campagnolo *et al.*, 2007; Ito *et al.*, 2005). The predominant reasons for the difficulty in obtaining better crystals of Pax8 and Pax6 in the previous studies were proposed to be intrinsic disorder of the crystalline lattice and anisotropic diffraction, respectively (Campagnolo *et al.*, 2007; Ito *et al.*, 2005). In the future, we hope to perform isomorphous replacement with heavy atoms while also carrying out optimization experiments with a much broader range of DNA elements to improve the quality of the crystals and ultimately solve the structure of Pax9 PD with the DC5 enhancer element.

This work was supported by the Agency for Science, Technology and Research (A*STAR), Singapore. We thank the National Synchrotron Light Source (NSLS), Brookhaven for access to the synchrotron-radiation beamline X29A and Dr Howard Robinson for providing assistance in data collection. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine. RJ is supported by the People's Government of Guangzhou Municipality Science and Technology Project 2011Y2-00026.

References

Baburajendran, N., Jauch, R., Tan, C. Y., Narasimhan, K. & Kolatkar, P. R. (2011). *Nucleic Acids Res.* **39**, 8213–8222.
Balczarek, K. A., Lai, Z.-C. & Kumar, S. (1997). *Mol. Biol. Evol.* **14**, 829–842.

Campagnolo, M., Pesaresi, A., Zelezetsky, I., Geremia, S., Randaccio, L., Bisca, A. & Tell, G. (2007). *J. Biomol. Struct. Dyn.* **24**, 429–441.
Epstein, J., Cai, J., Glaser, T., Jepeal, L. & Maas, R. (1994). *J. Biol. Chem.* **269**, 8355–8361.
Garvie, C. W., Hagman, J. & Wolberger, C. (2001). *Mol. Cell*, **8**, 1267–1276.
Inoue, M., Kamachi, Y., Matsunami, H., Imada, K., Uchikawa, M. & Kondoh, H. (2007). *Genes Cells*, **12**, 1049–1061.
Ito, M., Oyama, T., Okazaki, K. & Morikawa, K. (2005). *Acta Cryst.* F**61**, 1009–1012.
Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. (2001). *Genes Dev.* **15**, 1272–1286.
Li-Kroeger, D., Cook, T. A. & Gebelein, B. (2012). *Development*, **139**, 1611–1619.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Ng, C. K. L., Palasingam, P., Venkatachalam, R., Baburajendran, N., Cheng, J., Jauch, R. & Kolatkar, P. R. (2008). *Acta Cryst.* F**64**, 1184–1187.
Noll, M. (1993). *Curr. Opin. Genet. Dev.* **3**, 595–605.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pellizzari, L., Tell, G. & Damante, G. (1999). *Biochem. J.* **337**, 253–262.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L. & Pabo, C. O. (1999). *Genes Dev.* **13**, 1263–1275.