

**Peng Lu, Jianguo Liu, Manana
 Melikishvili, Michael G. Fried
 and Young-In Chi***

Department of Molecular and Cellular
 Biochemistry and Center for Structural Biology,
 University of Kentucky, Lexington, KY 40536,
 USA

Correspondence e-mail: ychi@uky.edu

Received 29 January 2008

Accepted 14 March 2008

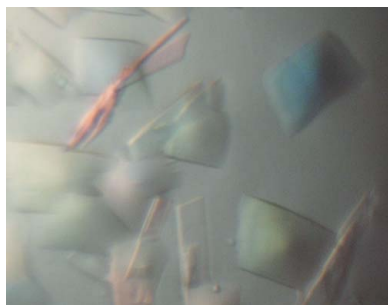
Crystallization of hepatocyte nuclear factor 4 α (HNF4 α) in complex with the HNF1 α promoter element

Hepatocyte nuclear factor 4 α (HNF4 α) is a member of the nuclear receptor superfamily that plays a central role in organ development and metabolic functions. Mutations on *HNF4 α* cause maturity-onset diabetes of the young (MODY), a dominant monogenic cause of diabetes. In order to understand the molecular mechanism of promoter recognition and the molecular basis of disease-causing mutations, the recombinant HNF4 α DNA-binding domain was prepared and used in a study of its binding properties and in crystallization with a 21-mer DNA fragment that contains the promoter element of another MODY gene, *HNF1 α* . The HNF4 α protein displays a cooperative and specific DNA-binding activity towards its target gene-recognition elements. Crystals of the complex diffract to 2.0 Å using a synchrotron-radiation source under cryogenic (100 K) conditions and belong to space group *C2*, with unit-cell parameters $a = 121.63$, $b = 35.43$, $c = 70.99$ Å, $\beta = 119.36^\circ$. A molecular-replacement solution has been obtained and structure refinement is in progress. This structure and the binding studies will provide the groundwork for detailed functional and biochemical studies of the MODY mutants.

1. Introduction

Hepatocyte nuclear factor 4 α (HNF4 α) is a tissue-specific transcription factor that plays an essential role in early vertebrate development and embryonic survival. It regulates the expression of a wide variety of essential genes, including those involved in liver and pancreatic cell differentiation (Li *et al.*, 2000; Odom *et al.*, 2004), embryogenesis and early development (Duncan *et al.*, 1994; Lausen *et al.*, 2000), glucose metabolism (Stoffel & Duncan, 1997), lipid homeostasis (Hayhurst *et al.*, 2001) and amino-acid metabolism (Schrem *et al.*, 2002). Mutations in *HNF4 α* cause a dominantly inherited form of diabetes known as maturity-onset diabetes of the young (MODY; Yamagata *et al.*, 1996). These mutations cause the loss of function of the gene product (Lausen *et al.*, 2000), which leads to impaired insulin secretion and defects in metabolic pathways (Miura *et al.*, 2006).

HNF4 α is a prototypical member of a unique nuclear receptor superfamily (NR2A1; Nuclear Receptors Nomenclature Committee, 1999) and exclusively functions as a homodimer (Jiang *et al.*, 1995), despite its sequence homology to retinoic X receptor (RXR), which can readily heterodimerize with a related nuclear receptor (Szanto *et al.*, 2004). HNF4 α consists of distinctive functional domains including a DNA-binding domain (DBD), a ligand-binding domain (LBD) and additional domains with transcription-activation functions (AF; Schrem *et al.*, 2002). However, the identity of its *bona fide* ligand is still under dispute (Hertz *et al.*, 1998; Petrescu *et al.*, 2005), even though its apparent ligand has been identified from structural studies (Dhe-Paganon *et al.*, 2002; Wisely *et al.*, 2002). HNF4 α -DBD contains two zinc-finger motifs that specifically recognize and bind as a homodimer to a direct repeat of two hexameric half-sites separated by one (DR1; in the majority) or two nucleotides (DR2) (Jiang *et al.*, 1995; Rajas *et al.*, 2002). Five MODY1 missense mutations (on four different residues) are found within the region of our HNF4 α -DBD



Healthcare). Data from serial dilution experiments were analyzed using the equations

$$\ln \frac{[P_n D]}{[D]} = n \ln [P] + \ln K_{\text{obs}} \quad (1)$$

and

$$[P] = [P]_{\text{tot}} - n[P_n D]. \quad (2)$$

Here, n is the binding stoichiometry, $[P_n D]$ and $[D]$ the concentrations of complex and free DNA, respectively, and $K_{\text{obs}} = [P_n D]/[P]^n[D]$. When both n and K_{obs} are unknown (as in this case), iterative calculation of n and $[P]$ using (1) and (2) results in convergence on self-consistent values of n and K_{obs} (Adams & Fried, 2007; Fried & Crothers, 1984).

2.5. Dynamic light-scattering measurement

The effective molecular radius and the homogeneity/monodispersity of the complex within various particular buffer conditions were measured using the Solubility Screening Kit (Jena Biosciences) in conjunction with a Dynapro-99 dynamic light-scattering instrument (Proterion Corporation) and a DynaPro-MSTC200 micro-sampler (Protein Solutions). The results were analyzed using *DYNAMICS* v.5.26.60 (Protein Solutions). 20 μl of sample was inserted into the cuvette with the temperature control set to 293 K. The light-scattering signal was collected at a wavelength of 830.7 nm. Protein concentrations were about 2 mg ml⁻¹ in each buffer and an average of 15 readings were recorded for each measurement.

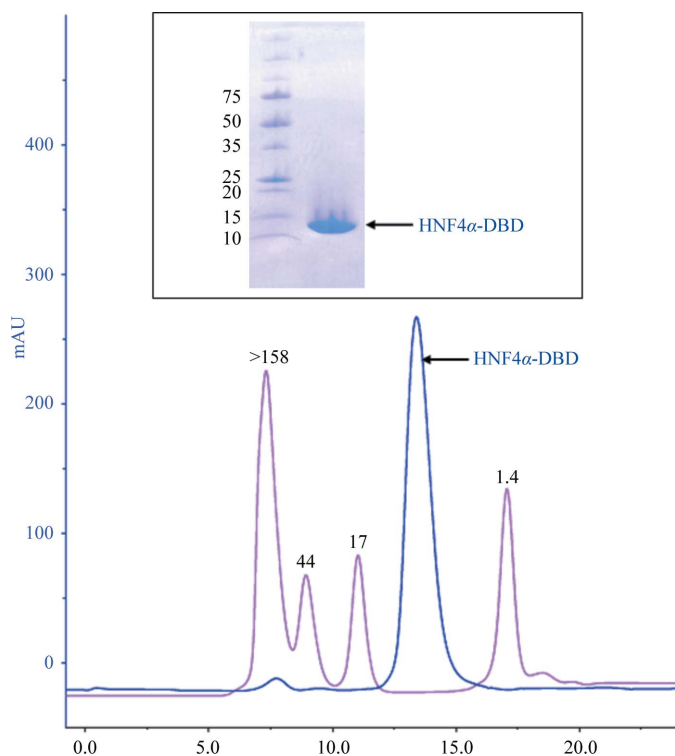


Figure 2 Gel filtration of HNF4 α -DBD with the molecular-weight standard samples (labelled in kDa). HNF4 α -DBD is homogeneous and appears to be a monomer in solution. The SDS-PAGE of purified HNF4 α -DBD along with the molecular-weight standard is shown in the inset.

2.6. Crystallization and optimization

Protein–DNA complexes were dialyzed in 20 mM Tris pH 7.5, 75 mM NaCl and 1 mM DTT at 277 K for 2.5 h and concentrated to at least 10 mg ml⁻¹. The initial crystallization trials were carried out at 295 K in 24-well plates using the hanging-drop vapor-diffusion method with a sparse-matrix screen (Jancarik & Kim, 1991) and similar commercially available versions such as Crystal Screens I and II (Hampton Research), Natrix and PEG/Ion Screens (Hampton Research), Cryo I and II (Molecular Dimensions) and Wizard I and II (deCODE Genetics).

Drops consisting of 0.5 μl protein–DNA solution were mixed with an equal volume of reservoir solution and equilibrated against 500 μl reservoir solution. Although many different DNA constructs varying

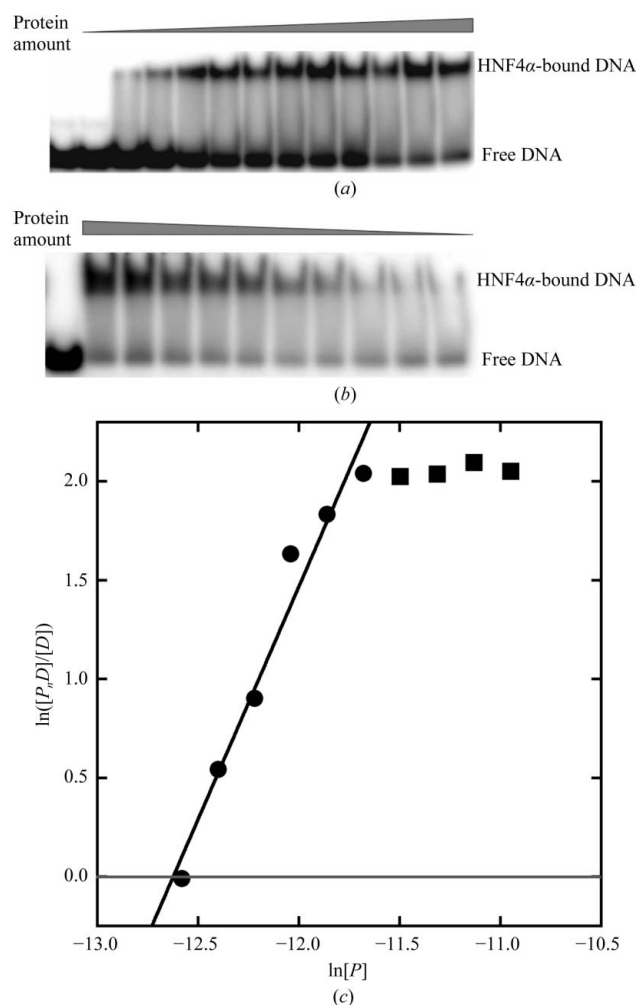


Figure 3 Binding of HNF4 α -DBD to duplex 21-mer target DNA detected by EMSA. Reactions were carried out in 10 mM Tris (pH 8.0 at 293 K), 1 mM EDTA, 100 mM NaCl, 1 mM MgCl₂, 1 mM DTT, 4% glycerol, 100 $\mu\text{g ml}^{-1}$ BSA. (a) Forward titration. All samples contained 0.21 μM duplex 21-mer. Samples contained HNF4 α -DBD protein, from the second lane, at 0.41, 0.82, 1.23, 2.05, 2.87, 3.69, 4.51, 5.33, 6.56, 8.2, 12.0 and 20.5 μM , respectively. (b) Serial dilution. The first lane contained reference DNA; the initial sample (second lane) contained 0.2 μM duplex 21-mer and 17.9 μM HNF4 α -DBD protein. Each succeeding lane contained an aliquot of the previous sample diluted 1.2-fold. (c) Determination of stoichiometry and association constant. Graph of $\ln([P_n D]/[D])$ as a function of $\ln[P]$. The slope about the mid-point of the reaction (where $\ln([P_n D]/[D]) = 0$) gives the stoichiometry ($n = 2.3 \pm 0.2$). The data used for this determination are indicated by filled circles. The data near binding saturation (filled squares) deviate from the linear relationship and were excluded from the fit. A fit of (1) to the linear data returned an intercept value of $\ln[P] = -12.58 \pm 0.03$, equivalent to $K_{\text{obs}} = 8.48 \pm 0.67 \times 10^{10} \text{ M}^{-2}$.

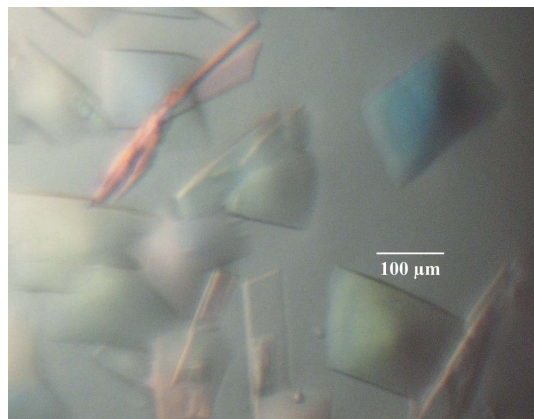


Figure 4
Typical crystals of the HNF4 α -DNA complex.

in length and the nature of the ending (blunt end *versus* overhang) were screened, diffraction-quality crystals were only reproducibly obtained using the overhang 21-mer shown in Fig. 1(b) (the two HNF4 α direct-repeat recognition sites are indicated by the boxes). Conditions yielding small crystals were further optimized by variation of the crystallization parameters and additives. The final condition, which produced somewhat flat bipyramidal crystals at 295 K, contained 26%(v/v) PEG 4000, 80 mM magnesium acetate and 50 mM sodium citrate pH 4.8.

2.7. Data collection and processing

The crystals were transferred into mother liquor containing an additional 15%(v/v) glycerol as a cryoprotectant before being directly plunged into liquid nitrogen and stored for data collection. The native data were collected at 100 K at APS (SER-CAT 22BM) using a MAR 225 CCD detector and an oscillation angle of 1° with 2 s exposure and

were processed using *HKL-2000* (Otwinowski & Minor, 1997). The wavelength used was 0.92017 Å.

3. Results and discussion

Recombinant HNF4 α -DBD (amino acids 46–126; Fig. 1a) was purified to homogeneity and mixed with pure DNA for subsequent studies. Gel-filtration experiments showed that the HNF4 α -DBD protein existed as a monomer in solution (Fig. 2). Purified HNF4 α -DBD protein forms a single complex with DNA containing its target sequence (Figs. 3a and 3b). Serial dilution analysis (Figs. 3b and 3c) revealed that the stoichiometry of the complex was 2:1 HNF4 α :dsDNA, with an association constant K_{obs} of $8.48 \pm 0.67 \times 10^{10} \text{ M}^{-2}$. The corresponding monomer equivalent dissociation constant was $3.43 \pm 0.13 \times 10^{-6} \text{ M}$. The formation of a 2:1 complex without the accumulation of detectable levels of the 1:1 intermediate indicates that binding is cooperative. These features will serve as a reference when we study the effects of MODY mutations on DNA binding in the near future.

Dynamic light scattering (DLS) is a useful tool to monitor protein solubility behavior and to predict favorable crystallization conditions (Wilson, 2003). We used the Solubility Screening Kit (Jena Biosciences) in conjunction with DLS (Jancarik *et al.*, 2004) in order to identify the optimal buffer conditions for complex formation and crystallization. The best polydispersity value of 0.06 was obtained with a buffer containing 20 mM Tris pH 7.5 and 75 mM NaCl and this optimal buffer was used for subsequent crystallization.

For crystallization, purified HNF4 α 46–126 and various DNAs were simply mixed in a 2:1.2 molar ratio, dialyzed against the optimal binding buffer (20 mM Tris pH 7.5, 75 mM NaCl) and concentrated using 10 kDa molecular-weight cutoff concentrators (Millipore). The protein-DNA concentration was 10 mg ml⁻¹ for initial screenings and 20 mg ml⁻¹ for final optimization. Crystals with the overhang 21-mer DNA (Fig. 1b) were grown at 295 K using the hanging-drop

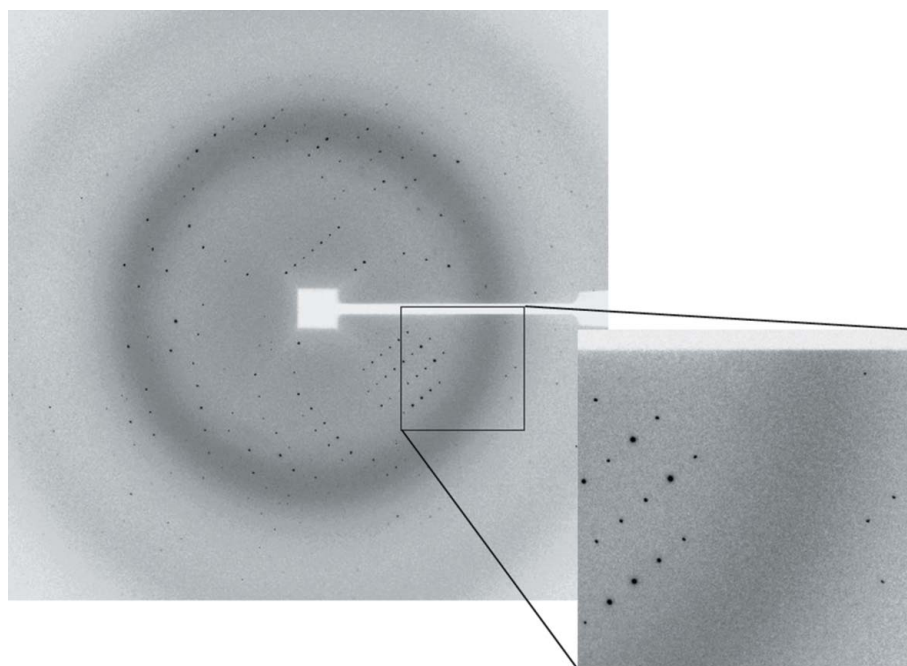


Figure 5
A typical X-ray diffraction pattern from a crystal of the HNF4 α -DNA complex. A small section near the water ring is enlarged and shown in the inset. The overall mosaicity of the crystal was 0.3°.

Table 1

Data-collection statistics for the HNF4 α -DNA crystal.

Values in parentheses are for the highest resolution bin.

| | |
|--|--|
| Space group | C2 |
| Unit-cell parameters (\AA , $^\circ$) | $a = 121.63$, $b = 35.43$, $c = 70.99$, $\beta = 119.36$ |
| Resolution (\AA) | 30.29–2.00 (2.07–2.00) |
| Observed reflections | 94972 |
| Unique reflections | 17094 |
| Redundancy | 5.6 (3.6) |
| Completeness (%) | 93.9 (72.4) |
| $I/\sigma(I)$ | 12.1 (4.5) |
| R_{merge}^\dagger (%) | 6.4 (30.4) |
| Matthews coefficient ($\text{\AA}^3 \text{Da}^{-1}$) | 2.12 |
| Solvent content (%) | 41.61 |
| Molecules per ASU | One complex (2 HNF4 α , 1 dsDNA) |

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the intensity of reflection hkl , \sum_{hkl} is the sum over all reflections and \sum_i is the sum over i measurements of reflection h .

vapor-diffusion method and the presence of the HNF4 α -DNA complex in the crystals was confirmed by running SDS-PAGE and 0.5% (w/v) agarose gels (data not shown). Crystals initially appeared within 2 d and continued to grow until they reached average dimensions of 0.05 \times 0.1 \times 0.2 mm (Fig. 4). A range of solution conditions varying the pH, temperature and concentrations of additives such as organic solvents, divalent cations and polyamines were used to attempt to improve the crystal quality. The final optimized condition contains 26% (v/v) PEG 4000, 80 mM magnesium acetate and 50 mM sodium citrate pH 4.8. The best crystal diffracted to 2.0 \AA at the synchrotron source and belongs to space group C2, with unit-cell parameters $a = 121.63$, $b = 35.43$, $c = 70.99 \text{\AA}$, $\beta = 119.36^\circ$ (Fig. 5). The value of the Matthews coefficient (Matthews, 1968) is 2.12 $\text{\AA}^3 \text{Da}^{-1}$ for one complex (two HNF4 α and one dsDNA) in the asymmetric unit and the estimated solvent content is 41.6% based on a protein specific density of 1.34. Final native data-collection statistics are summarized in Table 1.

The structure was determined by molecular replacement using the RXR-RAR-DNA complex structure (PDB code 1dsz) as a search model and the program *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP4* suite (Winn, 2003). An unambiguous solution was found that gave an initial R value of 51.4% and a correlation coefficient of 0.38 using data in the resolution range 15–3.0 \AA . The subsequent σ_A -weighted $2F_o - F_c$ map after rigid-body refinement clearly revealed density corresponding to the structural differences between the search model and the HNF4 α -DNA complex. Model improvement and refinement of the structure are in progress.

We wish to thank the staff at Advanced Photon Source beamline 22-BM (SER-CAT) for their help with data collection and David Rodgers for the use of the tissue-culture and protein-production core facility. Use of the Advanced Photon Source is supported by the US Department of Energy. This work was funded by a Juvenile Diabetes

Research Foundation grant (2004-503) and NIH Grant P20RR20171 from the COBRE program of the National Center for Research Resources to YIC and NIH grant GM070662 to MGF.

References

- Adams, C. & Fried, M. G. (2007). *Protein Interactions: Biophysical Approaches for the Study of Multicomponent Systems*, edited by P. Schuck, pp. 417–446. New York: Academic Press.
- Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y. I. & Shoelson, S. E. (2002). *J. Biol. Chem.* **277**, 37973–37976.
- Duncan, S. A., Manova, K., Chen, W. S., Hoodless, P., Weinstein, D. C., Bachvarova, R. F. & Darnell, J. E. Jr (1994). *Proc. Natl Acad. Sci. USA*, **91**, 7598–7602.
- Fried, M. G. & Crothers, D. M. (1984). *J. Mol. Biol.* **172**, 241–262.
- Gragnoletti, C., Lindner, T., Cockburn, B. N., Kaisaki, P. J., Gragnoli, F., Marozzi, G. & Bell, G. I. (1997). *Diabetes*, **46**, 1648–1651.
- Hayhurst, G. P., Lee, Y. H., Lambert, G., Ward, J. M. & Gonzalez, F. J. (2001). *Mol. Cell. Biol.* **21**, 1393–1403.
- Hellman, L. M. & Fried, M. G. (2007). *Nature Protocols*, **2**, 1849–1861.
- Hertz, R., Magenheimer, J., Berman, I. & Bar-Tana, J. (1998). *Nature (London)*, **392**, 512–516.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst. D60*, 1670–1673.
- Jiang, G., Nepomuceno, L., Hopkins, K. & Sladek, F. M. (1995). *Mol. Cell. Biol.* **15**, 5131–5143.
- Kritis, A. A., Argyrokastritis, A., Moschonas, N. K., Power, S., Katrakili, N., Zannis, V. I., Cereghini, S. & Talianidis, I. (1996). *Gene*, **173**, 275–280.
- Lausen, J., Thomas, H., Lemm, I., Bulman, M., Borgschulze, M., Lingott, A., Hattersley, A. T. & Ryffel, G. U. (2000). *Nucleic Acids Res.* **28**, 430–437.
- Li, J., Ning, G. & Duncan, S. A. (2000). *Genes Dev.* **14**, 464–474.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Maxam, A. M. & Gilbert, W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 560–564.
- Miura, A., Yamagata, K., Kakei, M., Hatakeyama, H., Takahashi, N., Fukui, K., Nammo, T., Yoneda, K., Inoue, Y., Sladek, F. M., Magnuson, M. A., Kasai, H., Miyagawa, J., Gonzalez, F. J. & Shimomura, I. (2006). *J. Biol. Chem.* **281**, 5246–5257.
- Nuclear Receptors Nomenclature Committee (1999). *Cell*, **97**, 161–163.
- Odum, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., Fraenkel, E., Bell, G. I. & Young, R. A. (2004). *Science*, **303**, 1378–1381.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Petrescu, A. D., Hertz, R., Bar-Tana, J., Schroeder, F. & Kier, A. B. (2005). *J. Biol. Chem.* **280**, 16714–16727.
- Rajas, F., Gautier, A., Bady, I., Montano, S. & Mithieux, G. (2002). *J. Biol. Chem.* **277**, 15736–15744.
- Schrem, H., Klempnauer, J. & Borlak, J. (2002). *Pharmacol. Rev.* **54**, 129–158.
- Stoffel, M. & Duncan, S. A. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 13209–13214.
- Szanto, A., Narkar, V., Shen, Q., Uray, I. P., Davies, P. J. & Nagy, L. (2004). *Cell Death Differ.* **11**, Suppl. 2, S126–S143.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.
- Winn, M. D. (2003). *J. Synchrotron Rad.* **10**, 23–25.
- Wisely, G. B., Miller, A. B., Davis, R. G., Thornquest, A. D. Jr, Johnson, R., Spitzer, T., Seffler, A., Shearer, B., Moore, J. T., Willson, T. M. & Williams, S. P. (2002). *Structure*, **10**, 1225–1234.
- Yamagata, K., Furuta, H., Oda, N., Kaisaki, P. J., Menzel, S., Cox, N. J., Fajans, S. S., Signorini, S., Stoffel, M. & Bell, G. I. (1996). *Nature (London)*, **384**, 458–460.