Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of the xeroderma pigmentosum group F endonuclease from *Aeropyrum pernix*

John Lally,[a] Matthew Newman,[a] Judith Murray-Rust,[a,b] Andrew Fadden,[a] Yutaka Kawarabayasi[c] and Neil McDonald[a,b]*

[a]Structural Biology Laboratory, The London Research Institute, Cancer Research UK, 44 Lincolns Inn Fields, London WC2A 3PX, England, [b]Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England, and [c]National Institute of Technology and Evaluation, Shibuya, Tokyo 151-0066, Japan

Correspondence e-mail: mcdonald@cancer.org.uk

Received 3 June 2004
Accepted 12 July 2004

The xeroderma pigmentosa group F protein (XPF) is a founding member of a family of 3′-flap endonucleases that play an essential role in nucleotide-excision repair, DNA replication and recombination. The XPF gene has been cloned from *Aeropyrum pernix*, encoding a 254-residue protein (apXPF). Recombinant protein was produced in *Escherichia coli* and purified by three chromatographic steps. Three different crystal forms of apXPF were grown in trigonal, monoclinic and triclinic systems. The trigonal crystals diffracted to 2.8 Å and were grown in the presence of double-stranded DNA. Monoclinic crystals were grown without DNA and diffracted to 3.2 Å. Triclinic crystals were grown from a truncated apXPF protein lacking the tandem helix–hairpin–helix motifs and diffracted to 2.1 Å.

1. Introduction

Nucleotide-excision repair (NER) is a complex repair pathway able to detect and remove a variety of bulky DNA lesions such as UVB-induced photoproducts from the genome of many organisms (Lindahl & Wood, 1999; Araujo & Wood, 1999). NER in eukarya involves the coordinated recruitment of a large number of proteins to recognize, unravel and excise a short oligonucleotide bearing the DNA lesion prior to filling in the missing gap (Araujo & Wood, 1999). The two unrelated endonucleases XPF and XPG selectively cleave the damaged DNA strand either side of the lesion and this is facilitated by their different respective substrate polarities (Petit & Sancar, 1999). XPF endonucleases recognize and cleave double-strand/single-strand DNA junctions with a single-stranded 3′ overhang, whilst XPG prefers similar substrates but with 5′ overhangs (Sijbers *et al.*, 1996; de Laat *et al.*, 1998; Hohl *et al.*, 2003).

XPF endonucleases are found in both eukaryotes and archaea (Sgouros *et al.*, 1999; White, 2003). They have an endonuclease domain followed by two consecutive helix–hairpin–helix motifs that form an $(HhH)_2$ domain (Nishino *et al.*, 2003; Doherty *et al.*, 1996). The endonuclease domain contains the catalytic motif GDX_nERKX_3D related to type II endonucleases, while the $(HhH)_2$ domain has been shown to mediate dimerization and to exhibit a sequence-independent DNA-binding function (Nishino *et al.*, 2003; Shao & Grishin, 2000). Eukaryotic XPFs have an SF2-like helicase domain at the amino-terminus that apparently lacks essential catalytic residues for ATPase activity (Sgouros *et al.*, 1999). A similar ‘long’ form of XPF is present in most euryarchaea, although notably their respective helicase domains are predicted to exhibit ATPase activity (White, 2003; Sgouros *et al.*, 1999; Komori *et al.*, 2002). Crenarchaea have a ‘short’ form of XPF that lacks a helicase-like domain and whose catalytic activity is regulated by interaction with PCNA (Roberts *et al.*, 2003). All XPFs studied require divalent cations for nuclease activity and can form either heterodimers (eukaryotic) with a shorter but structurally related binding partner or homodimers (archaea) (Sijbers *et al.*, 1996; Nishino *et al.*, 2003).

To investigate XPF architecture and the basis for substrate recognition, we initiated structural studies on a short XPF from *Aeropyrum pernix* (apXPF) that is 254 residues long (predicted molecular weight 28 726 Da). Here, we report the crystallization of three different crystal forms of apXPF from two truncated gene products corresponding to residues 19–231 (apXPF-ΔNΔC) and residues 19–150 (apXPF-ΔHhH_2). One crystal form was grown in the presence of double-stranded DNA to mimic an enzyme–product complex.

2. Expression and purification of apXPF

The gene encoding *A. pernix* XPF (accession codes C72622 and APE1043) was amplified from a plasmid generated from the *A. pernix* genome sequencing project (see http://www.bio.nite.go.jp:8080/dogan). Synthetic oligonucleotide primers were designed to remove regions of the protein that we anticipated to be flexible and that could hinder crystallization. The forward primer was 5′-TATAGGATCC-GGGTGGTCGTCCGCGTGTTTATGTGGA-TGTTAGGGAGGAG and the reverse primer

was 5′-TATAGGATCCCTATTATTAGCTACGCTTGTAAGGTGTCATGAGTATCTT. This deleted the C-terminal 23 residues that comprise the PCNA-binding motif and the N-terminal 18 residues (and will be referred to subsequently as apXPF-ΔNΔC). The oligonucleotides introduced a *Bam*H1 restriction site at either side of the amplified gene to allow subcloning into a modified expression vector (designated pET-14b-3C). This vector was derived from the pET-14b (Novagen) vector and altered to incorporate a 3C precision protease-cleavage site between the histidine tag and the cloned gene. This vector was transformed into Rosetta (DE3) pLysS bacterial cells (Novagen) for expression purposes.

A shorter construct was designed with a similar strategy and eliminated the $(HhH)_2$ domain (leaving residues 19–150, which we define as apXPF-ΔHhH_2). The oligonucleotide primers were 5′-TATAGGATCCGGGTGGTCGTCCGCGTGTTTATGTGGATGTTAGGGAGGAG (forward primer) and 5′-TATAGGATCCCTATTATTACTCTCTAGTGGAGAGGCGGGCGAGGCT (reverse primer). DNA sequencing confirmed that the correct sequence was obtained by PCR.

Both apXPF-ΔNΔC and apXPF-ΔHhH_2 proteins were expressed by induction with 0.5 m*M* IPTG overnight at 303 K when the Rosetta(DE3) pLysS cells reached an OD of 0.6. The apXPF-ΔNΔC protein was extracted following sonication of the cell pellets using a buffer comprised of 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 2 m*M* 2-mercaptoethanol (buffer *A*) and containing 5 m*M* imidazole, 1 m*M* PMSF, 10 m*M* benzamidine and 0.1 mg ml^{-1} DNAse. Clarified supernatant was mixed with Ni–NTA resin washed extensively in buffer A containing 20 m*M* imidazole. The polyhistidine-tagged apXPF-ΔNΔC was eluted with buffer *A* containing 0.5 *M* imidazole by pelleting the Ni–NTA resin for 2 min at 500*g*. The eluted protein was then dialysed overnight against buffer *A*. The polyhistidine tag was eliminated by addition of 100 µg of 3C protease per 10 mg recombinant protein at 277 K overnight, followed by the addition of a small amount of Ni–NTA agarose to remove the tag and uncleaved protein. Purified protein was then concentrated to approximately 4 mg ml^{-1} using Centriprep10 (Amicon Corporation) prior to crystallization. Recombinant apXPF-ΔHhH_2 protein was extracted and purified in an identical manner to the longer construct except that all buffers included 0.5%(*w*/*v*) CHAPS (Sigma). The detergent was needed in order to efficiently elute the protein from the Ni–NTA resin. Both recombinant apXPF-ΔNΔC and apXPF-ΔHhH_2 proteins were biosynthetically labelled with selenomethionine using the methionine-auxotrophic bacterial strain B834 (Novagen). These cells were transformed with either of the pET14b-apXPF constructs together with the Rosetta pLysS plasmid and were grown in minimal media supplemented with selenomethionine using standard techniques (Doublié, 1997).

Following Ni–NTA agarose purification, native and selenomethionine-labelled apXPF-ΔNΔC proteins were further purified by a single-stranded DNA (ssDNA) agarose column and size-exclusion chromatography using an identical protocol. The ssDNA affinity column was prepared using denatured calf thymus DNA agarose according to the manufacturers' instructions (Amersham-Pharmacia). The apXPF-ΔNΔC protein was dialysed into 20 m*M* Tris pH 8.0, 0.05 *M* NaCl and loaded onto the affinity column. A linear gradient of 0.05–1 *M* NaCl was applied using an ÄKTA FPLC system. The bound protein eluted as a single peak at 0.6 *M* NaCl. This sample was then loaded onto a Pharmacia Superdex 75 column pre-equilibrated in 20 m*M* Tris pH 8.0, 0.5 *M* NaCl, 1 m*M* DTT. Native apXPF-ΔNΔC has a predicted molecular weight of 25 097 Da. Both native and selenomethionine-labelled apXPF-ΔNΔC proteins both migrated with an apparent molecular weight of 42 kDa, indicating that they are associated as dimers. We then conducted static light-scattering experiments on purified native apXPF-ΔNΔC proteins at 0.5 mg ml^{-1} in 20 m*M* Tris pH 8.0, 0.5 *M* NaCl, 1 m*M* DTT using a MiniDawn Instrument (Wyatt Technologies). These experiments showed that the protein was monodisperse and solving the Zimm equation (Wilson, 2003) gave an approximate molecular weight of 51.6 ± 0.5 kDa, indicating a dimer. Purified native and selenomethionine-labelled apXPF-ΔHhH_2 was also dimeric as judged by size-exclusion chromatography under similar buffer conditions [except for the inclusion of 0.5%(*w*/*v*) CHAPS in the buffer]. These data are consistent with studies on the related 'long' XPF from *Pyrococcus furiosus*, which was shown to be dimeric (Nishino *et al.*, 2003). This may be a more general property of archaeal XPFs and contrasts with the heterodimeric state eukaryotic XPF complexes (White, 2003).

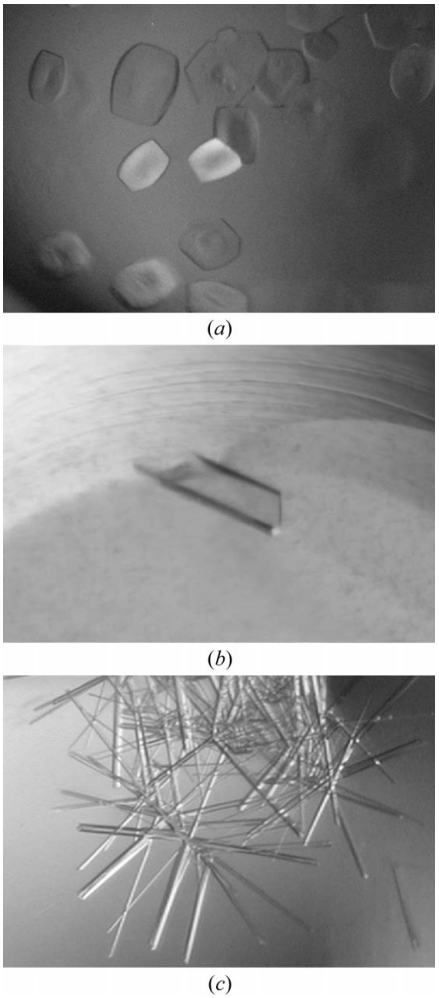

Figure 1
(*a*) Trigonal crystals of apXPF-ΔNΔC grown in the presence of a 15-mer dsDNA. (*b*) Triclinic crystals of apXPF-ΔHhH_2. (*c*) Monoclinic crystals of apXPF-ΔNΔC.

3. Crystallization

Purified apXPF-ΔNΔC was concentrated to 7 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 10 m*M* NDSB201 (Calbiochem) prior to crystallization. Needle-shaped crystals (Fig. 1) grew in 65%(*v*/*v*) MPD, 0.1 *M* Tris–HCl pH 8.0 and pH 9.0 in sitting drops and appeared after 6–10 weeks. Crystals of apXPF-ΔNΔC labelled with selenomethionine were also grown under similar conditions; however, their diffraction properties were very poor (typically ∼8–10 Å resolution) and they were not pursued further. To crystallize apXPF-ΔNΔC with dsDNA to mimic an enzyme–product complex, we prepared a series of different length duplex DNA ranging from 11 to 17 base pairs using complementary HPLC-purified oligonucleotides. These were mixed with 10 mg ml^{-1} protein stock solution in 20 m*M* Tris pH 8.0, 0.05 *M* NaCl, 2 m*M* 2-mercaptoethanol with a protein:DNA molar ratio of 1:3; this drop was then mixed 1:1 with well

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	apXPF-ΔHhH$_2$	apXPF-ΔNΔC	apXPF-ΔNΔC
Form	SeMet	Wild type	Wild type + dsDNA
Wavelength (Å)	0.9790	1.0000	0.9780
X-ray source	BM14, ESRF	Station 9.6, SRS	Station 9.6, SRS
Space group	*P*1	*C*2	$P3_221$ or $P3_121$
Unit-cell parameters			
a (Å)	33.6	210	141.3
b (Å)	38.7	42.7	141.3
c (Å)	55.3	118.7	85.3
α (°)	89.1		
β (°)	102.9	121.4	
γ (°)	115.7		
No. measurements	29154 (4370)	52273 (7746)	125841 (8403)
No. unique reflections	8080 (1196)	14887 (2159)	21619 (2463)
Resolution limit (Å)	2.5 (2.64–2.5)	3.2 (3.37–3.2)	2.8 (2.95–2.8)
Completeness	91.9 (88.7)	99.9 (100)	97.5 (92.8)
R_{sym}	7.6 (10.3)	6.2 (33.8)	8.3 (39.9)
$R_{anomalous}$	6.0 (8.5)	—	—
Average $I/\sigma(I)$	8.0 (6.2)	10.6 (2.4)	13.7 (2.8)

solution. The best quality crystals grew using a 15-mer dsDNA with sequence TCAG-CATCTGTGATC annealed to a complementary oligonucleotide. These barrel-shaped crystals (Fig. 1) grew in Hampton Crystal Screen II condition No. 13 [0.1 *M* sodium acetate, 0.2 *M* ammonium sulfate, 20%(*w*/*v*) PEG MME 2K] by the vapour-diffusion method using hanging drops at room temperature. Crystals took at least six weeks to grow. Similar crystals of selenomethionyl-labelled apXPF-ΔNΔC were grown in the presence of dsDNA but gave little or no diffraction and were not pursued further.

Native and selenomethionine-labelled apXPF-ΔHhH$_2$ proteins were concentrated to 3 mg ml^{-1} in 20 m*M* Tris pH 8.0, 0.5 *M* NaCl, 0.5%(*w*/*v*) CHAPS for crystallization. Crystals of both forms of apXPF-ΔHhH$_2$ were grown in 10%(*w*/*v*) PEG 400, 0.05 *M* sodium acetate pH 5.0 by mixing protein: well solution in a 1:1 ratio under mineral oil in hydrophobic vapour batch trays (Douglas Instruments). Crystals of native and selenomethionine-labelled proteins grew in a few days and were prepared using a Douglas Instruments IMPAX I-5 crystallization robot with 2 µl drops.

4. X-ray analysis

For data collection, the trigonal and triclinic crystals were transferred into Paratone-N (Hampton) for 1–2 min and then flash-cooled in a nitrogen stream at 120 K from an Oxford Instruments Cryostream. The monoclinic crystals of apXPF-ΔNΔC were flash-cooled directly from the drop, as the concentration of MPD used as precipitant was sufficiently high to act as a cryoprotectant. Data were collected on either station 9.6 at the SRS synchrotron source, Daresbury or on beamline BM14 at the ESRF, Grenoble. Data were processed and integrated using either *HKL* or *MOSFLM* data-processing software (Collaborative Computational Project, Number 4, 1994; Otwinowski & Minor, 1997) and space-group assignments were made based on auto-indexing solutions and consideration of systematic lattice absences. Data to 2.8 Å were measured from trigonal crystals of native apXPF-ΔNΔC grown in the presence of dsDNA. These crystals belong to space group $P3_121$ or $P3_221$ (Table 1). Monoclinic crystals of native apXPF-ΔNΔC belong to space group *C*2 and have a diffraction limit of 3.2 Å (Table 1). Crystals of apXPF-ΔHhH$_2$ were triclinic and therefore belong to space group *P*1. Although crystals of the native protein diffract to 2.1 Å, at this time we have only measured data to 2.5 Å for crystals of selenomethionine-labelled apXPF-ΔHhH$_2$.

To identify non-crystallographic twofold axes in each of the crystal forms, we inspected the $\chi = 180°$ section of a self-rotation function calculated using *MOLREP* (Vagin & Teplyakov, 2000). Using data from the triclinic crystals of apXPF-ΔHhH$_2$, a single non-crystallographic twofold axis at ($\theta = 90$, $\varphi = 0$, $\chi = 180°$) was evident with a peak height of 14.71σ parallel to *a*. This agrees with solvent calculations, which suggested a dimer in the asymmetric unit with a solvent content of 34.5%. Self-rotation function analysis for the monoclinic crystals of apXPF-ΔNΔC also revealed a non-crystallographic twofold axis at ($\theta = 28$, $\varphi = 0$, $\chi = 180°$) perpendicular to *b* with a peak height of 7σ. In this crystal form the asymmetric unit contents are less clear, as either one dimer (solvent content 71.7%) or two dimers (solvent content 43.4%) per asymmetric unit are feasible. The trigonal apXPF-ΔNΔC-dsDNA crystals may contain one dimer (solvent content 77.3%) or two dimers (solvent content 54.6%) in the asymmetric unit. However, we found no evidence of a non-crystallographic twofold axis over a range of resolution cutoffs and Patterson integration radii.

5. Discussion

The XPF gene has been cloned from *A. pernix* and expressed in a bacterial host to produce sufficient quantities of recombinant protein to initiate a structural analysis. Purified apXPF binds tightly to an ssDNA affinity column and we show from gel-filtration and light-scattering experiments that it is dimeric in solution. We report here three crystal forms of apXPF that diffract to moderate resolution at a synchrotron source. A self-rotation analysis of the trigonal crystal form showed no evidence of a non-crystallographic twofold. However, in high-symmetry space groups it is frequently observed that non-crystallographic and crystallographic symmetries coincide, therefore effectively masking the presence of the former. In contrast, self-rotation searches on the monoclinic and triclinic crystal forms showed clear evidence of a non-crystallographic twofold axis consistent with the solution data.

To obtain phasing information, we prepared crystals of selenomethionine-labelled protein. However, this phasing strategy was hindered by the lack of diffraction from crystals of selenomethionine-labelled apXPF-ΔNΔC. We therefore purified and crystallized a shorter form of apXPF lacking the (HhH)$_2$ domain. Crystals of selenomethionine-labelled apXPF-ΔHhH$_2$ diffract as well as native apXPF-ΔHhH$_2$ crystals and should provide experimental phasing for the endonuclease domain. While this work was being undertaken, the structure of the Hef/pfXPF nuclease domain was determined (Nishino *et al.*, 2003). The availability of these coordinates (50% identity to apXPF) gives an alternative route to phasing data from apXPF-ΔHhH$_2$ crystals by molecular replacement. An apXPF endonuclease domain structure will provide a partial model to phase data from crystals of apXPF-ΔNΔC. Although molecular-replacement search models are available for the apXPF (HhH)$_2$ domain, these are quite divergent in both sequence and structure (Shao & Grishin, 2000). The closest to apXPF is the UvrC (HhH)$_2$ domain, which is only 30%

identical and was determined by NMR (Singh *et al.*, 2002). We therefore intend to supplement phasing information for the apXPF-ΔNΔC crystals by screening heavy-atom derivatives and/or using iodinated oligonucleotides. The crystals described here provide an opportunity to determine the first structure of a full-length XPF and to understand how XPF recognizes and cleaves ds-ssDNA junction substrates.

This work was supported by Cancer Research UK. We gratefully acknowledge discussions with Professor Rick Wood during the early stages of this project.

References

Araujo, S. J. & Wood, R. D. (1999). *Mutat. Res.* **435**, 23–33.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–764.
Doherty, A. J., Serpell, L. C. & Ponting, C. P. (1996). *Nucleic Acids Res.* **24**, 2488–2497.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Hohl, M., Thorel, F., Clarkson, S. G. & Scharer, O. D. (2003). *J. Biol. Chem.* **278**, 19500–19508.
Komori, K., Fujikane, R., Shinagawa, H. & Ishino, Y. (2002). *Genes Genet. Syst.* **77**, 227–241.
Laat, W. L. de, Appeldoorn, E., Jaspers, N. G. & Hoeijmakers, J. H. (1998). *J. Biol. Chem.* **273**, 7835–7842.
Lindahl, T. & Wood, R. D. (1999). *Science*, **286**, 1897–1905.
Nishino, T., Komori, K., Ishino, Y. & Morikawa, K. (2003). *Structure*, **11**, 445–457.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Petit, C. & Sancar, A. (1999). *Biochimie*, **81**, 15–25.
Roberts, J. A., Bell, S. D. & White, M. F. (2003). *Mol. Microbiol.* **48**, 361–371.
Sgouros, J., Gaillard, P. H. & Wood, R. D. (1999). *Trends Biochem. Sci.* **24**, 95–97.
Shao, X. & Grishin, N. V. (2000). *Nucleic Acids Res.* **28**, 2643–2650.
Sijbers, A. M., de Laat, W. L., Arisa, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H. & Wood, R. D. (1996). *Cell*, **86**, 811–822.
Singh, S., Folkers, G. E., Bonvin, A. M., Boelens, R., Wechselberger, R., Niztayev, A. & Kaptein, R. (2002). *EMBO J.* **21**, 6257–6266.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
White, M. F. (2003). *Biochem. Soc. Trans.* **31**, 690–693.
Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.