*Acta Cryst.* (1998). D**54**, 1014–1016

# Crystallographic characterization of Pap1–DNA complex

YOSHIFUMI FUJII,[a] TAKESHI OHIRA,[b] YOSHIMASA KYOUGOKU,[c] TAKASHI TODA,[d] MITUHIRO YANAGIDA[e] AND TOSHIO HAKOSHIMA[a]* *at* [a]*Department of Molecular Biology, Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma, Nara 630-01, Japan,* [b]*Faculty of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565, Japan,* [c]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565, Japan,* [d]*Laboratory of Cell Regulation, Imperial Cancer Research Fund, PO Box 123, 44 Lincoln's Inn Fields, London WC2A 3PX, England, and* [e]*Department of Biophysics, Faculty of Science, Kyoto University, Kitashirakawaoiwake, Sakyo-ku, Kyoto 606-01, Japan. E-mail: hakosima@bs.aist-nara.ac.jp*

(*Received 15 December 1997; accepted 23 February 1998*)


## Abstract

Pap1 is a fission yeast transcription factor that activates genes related with resistance against staurosporine, a potent inhibitor of protein kinase C, and has been shown to be involved in cell growth, cell cycle, carcinogenesis and differentiation. Pap1 has the bZIP DNA-binding domain but binds to non-consensus DNA sequences for the bZIP motif. Highly ordered crystals of the DNA-binding domain complexed with a DNA fragment that has an ATF/CREB-like non-consensus sequence have been obtained. The crystals grew by the vapor-diffusion technique with polyethylene glycol 6000 and belong to space group $R3$ with $a = b = 240.78$, $c = 43.85$ Å. A 2.0 Å resolution data set was collected with a cryo-crystallographic technique.


## 1. Introduction

A fission yeast *Shizosaccaromyces pombe* gene, *pap1*+, encodes a transcription factor Pap1 that activates genes related with resistance against staurosporine, a potent inhibitor of protein kinase C (Toda *et al.*, 1991). Pap1, consisting of 544 amino-acid residues, contains a region of residues 71–140 that display similarity to the basic region followed by leucine-zipper DNA-binding motif (bZIP), which has been found in a superfamily of eukaryotic transcription factors such as the mammalian transcription factor AP-1. The bZIP transcription factors are dimerized through the leucine-zipper regions and recognize their DNA sequences through the basic regions that are rich in basic residues. In some cases, the leucine-zipper regions mediate hetero-dimerization with other bZIP transcription factors. The bZIP transcription factors form one of the largest superfamilies of eukaryotic transcription factors, which are engaged in cell growth, cell cycle, carcinogenesis and differentiation.

Pap1, an *S. pombe* AP-1-like transcription factor, forms a homodimer to bind to its unique DNA sequences. It has been demonstrated that, in general, two closely related types of consensus DNA sequence are the major target sites for bZIP transcription factors. The sequence **ATGAGTCAT** (the most strongly conserved palindromic bases are in bold) is referred to as the AP-1 site, since it is recognized by the mammalian AP-1 factor (Angel *et al.*, 1988) as well as by the Jun/Fos heterodimer (Rauscher *et al.*, 1988) and yeast GC4 (Hill *et al.*, 1986). The ATF/CREB sequence, **ATGACGTCAT**, is bound by CREB (Dwarki *et al.*, 1990), HBP-1 (Mikami *et al.*, 1989), CRE-BP1 (Maekawa *et al.*, 1989), and some of the transcription factors of the ATF family (Hai *et al.*, 1989). The ATF/CREB site is palindromic,

5′-**ATGA**C**GTCAT**
**TACTG**C**AGTA**-5′,

and the AP-1 site is pseudo palindromic with a pseudo dyad axis at the central GC base pair as follows.

5′-**ATGA**G**TCAT**
**TACTC**A**GTA**-5′.

More divergent sequences, however, are known to be recognized by other members of the bZIP family, including Pap1: ATF/CREB-like sequences. T**TTA**CG**TAA**T for Pap1 (Toda *et al.*, 1992) and A**TTG**CG**CAA**T for C/EBP (Landschulz *et al.*, 1989). Furthermore, non-palindromic sequences are also recognized by the bZIP transcription factors: A**TTAGTCAG**, which may be an AP-1-like sequence, for Pap1 and CACGTGGC for TAF-1 (Oeda *et al.*, 1991) and EmBP-1 (Guiltinan *et al.*, 1990). The same bZIP transcription factors can bind to a second DNA sequences with relatively high affinity, as well as Pap1 that binds to both non-consensus ATF/CREB-like and AP-1-like sites as described above. It has been reported that GCN4 and Jun/Fos also bind to the ATF/CREB site (Nakabeppu & Nathans, 1989; Sellers *et al.*, 1990) and that TAF-1 and EmBP-1 bind to an ATF/CREB-like site.

The crystal structures of GCN4 bound to the ATF/CREB site (Ellenberger *et al.*, 1992; König & Richmond, 1993) and Jun/Fos heterodimer bound to the AP-1 site (Glover & Harrison, 1995) have revealed the fundamental framework of the recognition of the DNA sequences by the bZIP motifs. Nevertheless, the question of how the bZIP motif recognizes the diverged DNA sequences, and particularly in the case of those having non-consensus sequences remains to be clarified. The basic region of Pap1 exhibits a high homology (60%) with the counterpart in budding yeast, YAP1, but relatively low homologies (33–37%) with other bZIP transcription factors, such as c-Jun, c-Fos, and GCN4. Therefore, the recognition of the non-consensus sequences by Pap1 must be an intrinsic property of its basic region. Moreover, the details of the protein–DNA interactions involving water molecules remain unclear due to the limitation of the resolutions (3 Å) at which the structures of the GCN4–DNA and Jun/Fos–DNA complexes were determined, although recently, one of the GCN4 structures was refined at 2.2 Å resolution (Keller *et al.*, 1995). To clarify these points, we obtained high-resolution crystals of the bZIP domain of Pap1 with a DNA oligomer containing the non-consensus ATF-/CREB-like sequence.

## 2. Materials and methods

The DNA-binding domain of Pap1, consisting of 70 residues, was expressed in BL21(DE3) using a T7 expression system (Studier & Moffatt, 1986). The DNA-binding domain restores the DNA-binding activity (Toda *et al.*, 1992). The construction and expression of the plasmids were as described previously (Kyogoku *et al.*, 1993). The cells were suspended in 50 m*M* potassium phosphate buffer (pH 6.8) containing 50 m*M* KCl, 1 m*M* $NaN_3$, 1 m*M* phenylmethane sulfonyl fluoride, and 7 m*M* 2-mercaptoethanol and were disrupted by sonication at 277 K. The supernatant was treated with 20% polyethyleneimine to remove nucleic acids and acidic proteins. The supernatant was fractionated with ammonium sulfate. The fractions with between 45 and 70% saturated ammonium sulfate were collected for further purification by two column-chromatographic steps, using SP-Sepharose and Mono-S (Pharmacia Biotech). The purity was monitored by 17.5% polyacrylamide gel electrophoresis with staining by Coomassie Brilliant Blue. The DNA oligomers used in the crystallization attempts were synthesized by standard phosphoramidite chemistry and purified by ion-exchange, gel-filtration and reverse-phase HPLC column chromatography, using Resource Q (Pharmacia Biotech) Hitrap Desalting (Pharmacia Biotech) and TSK-gel ODS80Ts (Tosoh, Japan), respectively. The purity was monitored by a C-18 reverse-phase HPLC column chromatography using Sensyu Pak ODS-1251-N (Sensyukagaku, Japan). The determination of the concentration of each oligomers and the annealing of the strands to form double-stranded DNA oligomers were as described previously (Hakoshima *et al.*, 1993). The sequences of the DNA oligomers were based on the non-consensus sequences from the Pap1 binding sites of the *p25* gene (Toda *et al.*, 1992).

Crystallization conditions were screened using the sitting-drop and hanging-drop vapor-diffusion methods with various precipitants over a wide pH range. The chain lengths of DNA oligomers were varied systematically from 9 to 18 nucleotides, and both duplexes with blunt ends and with unpaired bases were used in attempts to obtain highly ordered crystals. We also tried to evaluate detergents for their ability to influence the crystallization, using a commercial detergent screening kit (Hampton Research, USA). The crystals obtained were resolved in an aliquot of a buffer for electrophoretic mobility shift assay with 6% acrylamide gels to verify that the crystal contains the protein–DNA complexes. X-ray diffraction data of the crystal were collected with a Rigaku imaging-plate area detector (R-AXIS IV) using Cu $K\alpha$ radiation ($\lambda$ = 1.54178 Å) generated by a rotating anode (Rigaku FR-C) operating at 50 kV and 60 mA. The focus size of the X-ray beam was 100 µm, and the beam was focused using a Supper double-focusing mirror. The distance from crystal to imaging plate was 150–200 mm. The loop-mounted crystal was flash-cooled using a Rigaku cooling device with liquid nitrogen gas. Prior to the flash-cooling, crystals were transferred to solutions containing the mother liquor and cryo-protectants. The concentrations of cryo-protectants were gradually raised to 25% ethylene glycol and 10% polyethylene glycol 6000 (PEG 6K). Intensities were evaluated with the program *PROCESS* (Rigaku) (Higashi, 1990; Sato *et al.*, 1992). The self-rotation functions (Rossmann & Blow, 1962) were calculated using the program *POLARRFN* in the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Using the coordinates of GCN4 (2dgc), preliminary structure analyses by the molecular replacement methods were performed with the programs *AMoRe* (Navaza, 1994) and *X-PLOR* (DeLano & Brünger, 1995).

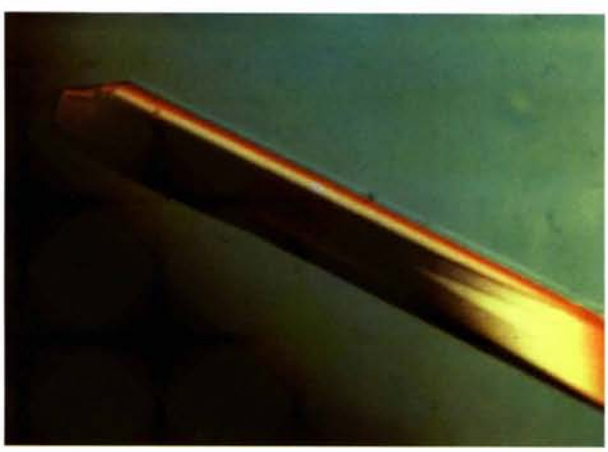

Fig. 1. Crystals of the DNA-binding domain of Pap1 and the cre-13s DNA oligomer. The longest one is 1 mm.

## 3. Results and discussion

For most DNA oligomers having chain lengths of 13–18 nucleotides, complex crystals appeared and grew to sizes suitable for X-ray experiments, whether the oligomers contain AP-1-like or ATF/CREB-like sequences. The crystals, however, diffracted to resolutions lower than 6 Å. Using 13-mer and 14-mer DNA oligomers having self-complementary sequences, highly ordered crystals, which diffract to higher than 3–4 Å resolution, were obtained. Among the DNA oligomers, a 13-mer DNA, AGG**TTA**CG**TAA**CC which is referred to as cre-13s, gave the best crystal, with diffraction up to 2.0 Å resolution. The crystals appeared in a solution containing 50 m*M* Mes-K buffer (pH 5.8), 0.36 m*M* Pap1, 0.36 m*M* cre-13s, 20 m*M* KCl, and 0.9% PEG 6K equilibrated against 2.5% PEG 6K, and grew to a maximum size of 1 × 0.2 × 0.1 mm (Fig. 1). The crystals belong to space group *R*3 with unit-cell dimensions $a = b = 240.78$, $c = 43.85$ Å. Intensity data were collected for up to 2.0 Å resolution. A total of 168 348 independent measurements were merged to obtain 58 094 unique reflections constituting 90.6% completeness with an $R_{\rm merge}$ of 8.05%. At the outmost shell between 2.25 and 2.0 Å, the completeness and the $R_{\rm merge}$ were 87.1 and 37.2%, respectively. Because the length of the *c* axis divided by 13 gives a value corresponding to the height of the one base-pair step of B-DNA structure, 3.37 Å, it is likely that the double helices of the cre-13s DNA oligomers orient along the *c* axis. The intensity distribution also supports this hypothesis.

The assumption of two, three and four 1:1 complexes in the asymmetric unit give the $V_m$ values (Matthews, 1968), 5.4, 3.6 and 2.7 Å $Da^{-1}$, respectively. These values are high compared to those of protein crystals that have a commonly found vlaue of 2.3 $Å^3$ $Da^{-1}$, and they correspond to solvent contents ranging from 56 to 78%. We calculated self-rotation functions (Rossmann & Blow, 1962) to look for non-crystallographic symmetry with several Patterson sphere radii. However, we failed to find signifcant peaks on the maps. Possibly, non-crystallographic symmetric axes may run along one of the

crystallographic axes, if present. As crystals of protein–DNA complexes occasionally have high values of $V_m$ with a solvent content higher than 70% (Toumoto *et al.*, 1997; Shimizu *et al.*, 1997), we could not rule out the possibiity that two or three complexes exist in the asymmetric unit. We also tried to solve the structure by molecular replacement using a search model based on the structure of the GCN4–DNA complex (König & Richmond, 1993). However, any unique solution was not obtained from preliminary searches with a polyalanine model for Pap1 bound to the DNA using different ranges of intensity data and integration radii. Structural analysis of the complexes using the multiple isomorphous replacement method with the introduction of 5-iodo-deoxyuridine or 5-iodo-deoxycytidine into the DNA oligomer, as well as analyses using molecular replacement methods, is in progress.

We thank Dr T. Shimizu and Mr M. Kato for technical support. This work was supported by Grants in Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan to TH (06276104, 07559010, 09308025).

## References

Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T. & Karin, M. (1988). *Nature (London)*, **332**, 166–171.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Dwarki, V. J., Montminy, M. & Verma, I. M. (1990). *EMBO J.* **9**, 225–232.
Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992). *Cell*, **71**, 1223–1237.
Glover, J. N. M. & Harrison, S. C. (1995). *Nature (London)*, **373**, 257–261.
Guiltinan, M. J., Marcotte, W. R. & Quatrano, R. S. (1990). *Science*, **250**, 267–271.
Hai, T., Liu, F., Coukos, W. J. & Green, M. R. (1989). *Genes Dev.* **3**, 2083–2090
Hakoshima, T., Teranishi, Y., Ohira, T., Suzuki, K., Shimizu, M., Kyogoku, Y., Ogawa, N. & Oshima, Y. (1993). *J. Mol. Biol.* **229**, 566–569.
Higashi, T. (1990). *J. Appl. Cryst.* **23**, 253–257.
Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986). *Science*, **234**, 451–457.
Keller, W., König, P. & Richmond, T. J. (1995). *J. Mol. Biol.* **254**, 657–667.
König, P. & Richmond, T. J. (1993). *J. Mol. Biol.* **233**, 139–154.
Kyogoku, Y., Kojima, C., Lee, S. J., Tochio, H., Sizuki, B., Matsuo, H. & Shirakawa, M. (1993). *Methods Enzymol.* **261**, 524–541.
Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989). *Science*, **243**, 1681–1688.
Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M. & Ishii, S. (1989). *EMBO J.* **8**, 2023–2028.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mikami, K., Takase, H., Tabata, T. & Iwabushi, M. (1989). *FEMS J.* **256**, 67–70.
Nakabeppu, Y. & Nathans, D. (1989). *EMBO J.* **8**, 3833–3841.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oeda, K., Salinas, J. & Chua, N.-H. (1991). *EMBO J.* **10**, 1793–1802.
Rauscher, F. J. III, Sambucetti, L. C., Curran, T., Distel, R. J. & Spiegelman, B. M. (1988). *Cell*, **52**, 471–480.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sato, M., Yamamoto, M., Imada, K., Katsube, Y., Tanaka, I. & Higashi, T. (1992). *J. Appl. Cryst.* **25**, 348–357.
Sellers, J. W., Vincent, A. C. & Struhl, K. (1990). *Mol. Cell. Biol.* **10**, 5077–5086.
Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y. & Hakoshima, T. (1997). *EMBO J.* **16**, 4689–4697.
Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.
Toda, T., Shimanuki, M., Saka, Y., Yamano, H., Adachi, Y., Sirakawa, M., Kyogoku, M. & Yanagida, M. (1992). *Mol. Cell Biol.* **12**, 5474–5484.
Toda, T., Shimanuki, M. & Yanagida, M. (1991). *Genes Dev.* **5**, 60–73.
Toumoto, A., Shimizu, T., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y. & Hakoshima, T. (1997). *Acta Cryst.* D**53**, 103–104.