Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic study of intein homing endonuclease II encoded in the archaeal DNA polymerase gene

Hiroshi Hashimoto,[a] Hitomi Takahashi,[a] Motomu Nishioka,[b] Shinsuke Fujiwara,[b] Masahiro Takagi,[b] Tadayuki Imanaka,[c] Tsuyoshi Inoue[a] and Yasushi Kai[a]*

[a]Department of Materials Chemistry, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan, [b]Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan, and [c]Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Yoshidahonmachi Sakyo-ku, Kyoto 606-8501, Japan

Correspondence e-mail: kai@chem.eng.osaka-u.ac.jp

Received 24 March 2000
Accepted 22 June 2000

Intein homing endonucleases are proteins spliced out from a precursor protein and site-specific enzymes that make double-strand breaks in inteinless alleles. Crystals of intein homing endonuclease II from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 (PI-*Pko*II) have been grown at room temperature using ammonium sulfate as a precipitant. The diffraction pattern of the crystal extends to 3.0 Å resolution at room temperature upon exposure to synchrotron X-rays at KEK-PF, Japan. The crystals have symmetry consistent with space group $C222_1$, with unit-cell parameters a = 107.6, b = 150.5, c = 146.8 Å. A full set of X-ray diffraction data were collected to 3.0 Å Bragg spacing from a native crystal with an overall $R_{\rm merge}$ of 4.8% and a completeness of 96.6%.

1. Introduction

Inteins are proteins excised from a precursor protein. The self-excision process, protein splicing, is a post-transitional process involving the precise excision of an intervening protein domain (termed an intein) and the ligation of the external N- and C-domains (termed exteins). The first intein discovered was the *Saccharomyces cerevisiae* VMA intein (PI-*Sce*I) intervening in the catalytic subunit of vacuolar H^+-ATPase (Kane *et al.*, 1990). Since then, inteins have been found in a variety of proteins in archaea, bacteria and single-cell eukarya. Although several inteins were identified experimentally, most of the recently described inteins were predicted from DNA sequences (reviewed by Perler *et al.*, 1997). For an updated list of inteins and their characteristics, refer to the Intein Database web site at http://www.neb.com/neb/inteins.html. All inteins share four conserved splice-junction residues: Ser, Thr or Cys at the intein N-terminus and a His–Asn pair at the intein C-terminus followed by Ser, Thr, Cys at the C-extein N-terminus. A mechanistic model for protein splicing has been proposed based on mutagenesis studies and biochemical analysis of the intermediates occurring in the splicing pathway (Xu & Perler, 1996).

Analyses of amino-acid sequences suggest that most inteins contain a LAGLIDADG motif characteristic of homing endonucleases, which were first discovered in mobile self-splicing introns. The endonuclease activities of several inteins were detected experimentally. Intein homing endonucleases cleave DNA within their recognition sequences to leave four base 3′-hydroxyl overhangs. The recognition sequences are generally asymmetrical and long, with sizes of 12–40 bp (Perler *et al.*, 1997). Intein homing endonuclease recognizes the DNA sequence lacking its coding DNA sequence (homing site) and cleaves the intein-less DNA to initiate intein homing.

Although many amino-acid sequences of inteins have been reported, crystal structures of only two inteins have been determined. One is the structure of PI-*Sce*I from *S. cerevisiae* (Duan *et al.*, 1997) and the other is GyrA intein (PI-*Mxe*I) found in the bacterial gyrase A subunit from *Mycobacterium xenopi* (Klabunde *et al.*, 1998). There is no structural information for archaeal inteins. The understanding of the protein-splicing process and the structural correlation of inteins is far from being systematic or sufficient based on three-dimensional structures.

The DNA polymerase gene (*polA*) from the hyperthermophilic archaeon *P. kodakaraensis* strain KOD1 codes the DNA polymerase precursor. The precursor protein contains mature DNA polymerase (KOD DNA polymerase) and two inteins, PI-*Pko*I and PI-*Pko*II, with molecular masses of 41 and 62 kDa, respectively (Takagi *et al.*, 1997). PI-*Pko*I and PI-*Pko*II have a conserved sequence for self-splicing at the N- (Cys and Ser, respectively) and C-terminus (His-Asn) and the LAGLIDADG motif of homing endonuclease. They cleave the DNA regions of the respective inteinless DNA sequences. The minimal recognition sequences for the intein homing endonucleases are 19 bp (5′-GATTTTAGATCCCTGTACC-3′) for PI-*Pko*I and

Table 1
Data-collection statistics for PI-*Pko*II.

Space group	$C222_1$
Unit-cell parameters (Å)	a = 107.6, b = 150.5, c = 146.8
Temperature (K)	293
Wavelength (Å)	1.00
Resolution limits (Å)	67.4–3.0
Measured reflections	107 723
Unique reflections	23 383
Completeness (%)	96.6
Mean $I/\sigma(I)$	21.5
$R_{\rm merge}$† (%)	4.8

† $R_{\rm merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \sum_i I(h)_i$, where $I(h)$ is the intensity of reflection h and $\sum_h$ and $\sum_i$ are the summations over all reflections and all measurements i.

16 bp (5′-CAGCTACTACGGTTAC-3′) for PI-*Pko*II (Nishioka *et al.*, 1998).

We are carrying out crystallographic studies on the products of the *polA* gene: mature KOD DNA polymerase, PI-*Pko*I and PI-*Pko*II. Structural analyses of these proteins are required in order to reveal structural information on the protein-splicing process, correlation between the extein and the inteins and the molecular evolution of inteins. We have previously reported a crystallographic study of KOD DNA polymerase (Hashimoto *et al.*, 1999). In this paper, we describe the crystallization and X-ray crystallographic study of PI-*Pko*II. The crystal structure of PI-*Pko*II will provide a novel insight based on the three-dimensional structure of homing endonuclease.

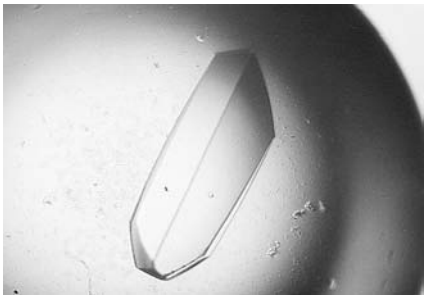

Figure 1
Crystal of intein homing endonuclease II from KOD1 (PI-*Pko*II).

2. Crystallization

PI-*Pko*II was overexpressed in *Escherichia coli* strain BL21(DE3) as previously reported (Nishioka *et al.*, 1998). Recombinant protein was purified by heat treatment, heparin agarose column chromatography (Sigma) and gel filtration (Superdex 75 HR10/30, Pharmacia). Purified PI-*Pko*II was concentrated up to about 20 Abs cm^{-1} at 280 nm in a 10 m*M* sodium phosphate buffer at pH 7.0 containing 0.1 m*M* EDTA, 1 m*M* DTT and 500 m*M* NaCl using Centricon-50 concentrators (Amicon). Crystals of PI-*Pko*II suitable for diffraction experiments were obtained at 293 K from hanging drops consisting of 2 µl of protein solution and 2 µl of reservoir solution (2.0 *M* ammonium sulfate) equilibrated against 500 µl of reservoir solution. The crystals reached maximum dimensions of 0.4 × 0.2 × 0.1 mm in several days (Fig. 1).

3. Data collection

Crystals of PI-*Pko*II were mounted in thin-walled glass capillaries with a small volume of reservoir solution. X-ray diffraction measurements were performed at beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba Science City, Japan. The wavelength used was 1.00 Å. The incident beam was collimated to 0.2 mm in diameter. Intensity data were collected on 400 × 800 mm imaging plates (Fuji Film Company Ltd) using the Weissenberg camera for macromolecules with a radius of 430 mm (Sakabe *et al.*, 1995; Watanabe *et al.*, 1995) and the oscillation method with 3° rotation per frame. The crystals diffracted at least to 3.0 Å resolution. The diffraction data were digitized with a Rigaku image analyzer. The digital data were then processed and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). A complete native data set was collected to 3.0 Å resolution using a single crystal of PI-*Pko*II. Table 1 shows a summary of the X-ray diffraction data statistics. The outermost resolution shell of data (3.11–3.00 Å) has an $R_{\rm merge}$ of 25.8% and a completeness of 96.4%.

We estimated the number of molecules in the asymmetric unit. Assuming two molecules in the asymmetric unit, the V_m value is calculated to be 2.4 Å^3 Da^{-1} (Matthews, 1968). Structure determination of PI-*Pko*II by multiple isomorphous replacement methods is under way.

We thank Professor N. Sakabe, Dr N. Watanabe, Dr M. Suzuki and Dr N. Igarashi for their support in data collection at KEK-PF, Japan. This study was supported by Sakabe Project TARA of the University of Tsukuba. HH is grateful to JSPS Fellowships for Japanese Junior Scientists.

References

Duan, X, Gimble, F. S. & Quiocho, F. A. (1997). *Cell*, **89**, 555–564.
Hashimoto, H., Matsumoto, T., Nishioka, M., Yuasa, T., Takeuchi, S., Inoue, T., Fujiwara, S., Takagi, M., Imanaka, T. & Kai, Y. (1999). *J. Biochem.* **125**, 983–986.
Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. & Stevens, T. H. (1990). *Science*, **250**, 651–657.
Klabunde, T., Sharma, S., Telenti, A., Jacobs W. R. Jr & Sacchettini, J. C. (1998). *Nature Struct. Biol.* **5**, 31–36.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nishioka, M., Fujiwara, S., Takagi, M. & Imanaka, T. (1998). *Nucleic Acids Res.* **26**, 4409–4412.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perler, F. B., Olsen, G. J. & Adam, E. (1997). *Nucleic Acids Res.* **25**, 1087–1093.
Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. & Sasaki, K. (1995). *Rev. Sci. Instrum.* **66**, 1276–1281.
Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). *Appl. Environ. Microbiol.* **63**, 4504–4510.
Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, N. (1995). *Rev. Sci. Instrum.* **66**, 1824–1826.
Xu, M.-Q. & Perler, F. B. (1996). *EMBO J.* **15**, 5146–5153.