

FOR THE RECORD

Crystallization and preliminary X-ray studies of I-*PpoI*: A nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*

KAREN E. FLICK,¹ DIANE McHUGH,¹ JOE DON HEATH,² KATHRYN M. STEPHENS,²
RAYMOND J. MONNAT, JR.,² AND BARRY L. STODDARD¹

¹Fred Hutchinson Cancer Research Center, 1124 Columbia St. A3-023, Seattle, Washington 98104

²University of Washington, Department of Pathology, Box 357705, Seattle, Washington 98195

(RECEIVED September 16, 1997; ACCEPTED October 10, 1997)

Abstract: The homing endonuclease I-*PpoI* is encoded by an optional third intron, Pp LSU 3, found in nuclear, extrachromosomal copies of the *Physarum polycephalum* 26S rRNA gene. This endonuclease promotes the lateral transfer or “homing” of its encoding intron by recognizing and cleaving a partially symmetric, 15 bp homing site in 26S rDNA alleles that lack the Pp LSU 3 intron. The open reading frame encoding I-*PpoI* has been subcloned, and the endonuclease has been overproduced in *E. coli*. Purified recombinant I-*PpoI* has been co-crystallized with a 21 bp homing site DNA duplex. The crystals belong to space group P3₁21, with unit cell dimensions $a = b = 114 \text{ \AA}$, $c = 89 \text{ \AA}$. The results of initial X-ray diffraction experiments indicate that the asymmetric unit contains an enzyme homodimer and one duplex DNA molecule, and that the unit cell has a specific volume of $3.4 \text{ \AA}^3/\text{dalton}$. These experiments also provide strong evidence that I-*PpoI* contains several bound zinc ions as part of its structure.

Keywords: endonuclease; intron homing; intron mobility; metal substitution; X-ray crystallography

Mobile intron endonucleases are a diverse family of proteins encoded by open reading frames contained within Group I self-splicing introns. Homologous endonucleases are also found as optional, independently folded domains in self-splicing protein introns termed “inteins.” The catalytic activity of these endonucleases promotes the lateral transfer of their encoding introns by a targeted transposition mechanism termed intron mobility or homing. Mobility or homing is initiated and targeted by the endonuclease. The enzyme makes a highly site-specific DNA double-strand break at a homing site that lacks the self-splicing intron (an

“intron-less” allele). The intron containing the endonuclease gene is subsequently transferred to the cleaved recipient allele by a double-strand break repair/gene conversion event. Mobile intron endonucleases have been identified in a diverse collection of lower eukaryotes, *Archea*, and eubacteria, with the largest number encoded in organellar genomes (Lambowitz & Belfort, 1993; Mueller et al., 1993; Belfort & Perlman, 1995; Belfort et al., 1995; Belfort & Roberts, 1997).

The largest family of mobile intron endonucleases contain members with one or two copies of the conserved LAGLIDADG or dodecapeptide protein motif (Lambowitz & Belfort, 1993; Belfort & Roberts, 1997). Proteins containing this motif include mobile intron endonucleases, self-splicing inteins (Cooper & Stevens, 1995), mitochondrial RNA maturases (Belfort & Perlman, 1995) and free-standing nuclear-encoded proteins such as the yeast HO mating-type switch endonuclease (Kostriken et al., 1983). The recently determined X-ray crystal structures of two mobile intron endonucleases, the *Chlamydomonas reinhardtii* I-CreI endonuclease and the yeast PI-*SceI* intein-encoded endonuclease, have emphasized the role of the LAGLIDADG motif in protein folding and the structure of the endonuclease active site (Duan et al., 1997; Heath et al., 1997). Several additional mobile intron endonuclease families have been identified that contain other conserved sequence motifs. These include the GIY-YIG, H-N-H and His-Cys box motifs (Belfort & Perlman, 1995; Belfort & Roberts, 1997).

The first identified member of the His-Cys box family was the I-*PpoI* endonuclease (Muscarella & Vogt, 1989; Muscarella et al., 1990). The I-*PpoI* endonuclease was identified as an open reading frame in the optional third intron, Pp LSU 3, found in nuclear, extra-chromosomal copies of the 26S rRNA gene of the acellular slime mold *Physarum polycephalum*. The Pp LSU 3 intron encoding I-*PpoI* is mobile in genetic crosses, and Pp LSU 3 transcripts can self-splice in vitro. The intron encoding I-*PpoI* is one of three mobile introns known to reside in a eukaryotic nucleus. The other two endonucleases were identified in nuclear introns of the slime mold *Didymium iridis* and the amoeba *Naegleria andersoni* ssp *andersoni* (Johansen et al., 1993; Johansen & Vogt, 1994).

Reprint requests to: Barry L. Stoddard, Fred Hutchinson Cancer Research Center, 1124 Columbia St. A3-023, Seattle, Washington 98104; e-mail: bstoddard@fred.hfcr.org.

Two protein products were predicted from the sequence of the Pp LSU 3 intron and adjacent 26S rRNA gene: a 138-residue "short" form of I-*PpoI* encoded entirely by the Pp LSU 3 intron, and a 160 residue "long" form of I-*PpoI* containing 22 additional N-terminal residues encoded by adjacent rDNA exon and intron sequences preceding the short form ATG at intron nucleotides 14–16 (Muscarella et al., 1990). The sequence of the intron was subsequently revised, altering I-*PpoI* short-form residues 122–125 and adding 35 residues to the C-terminus following residue 128 (E. Ellison, pers. obs.). Both of the predicted 163 and 185 residue I-*PpoI* proteins contain a single copy of the His-Cys box motif, and both are enzymatically active when synthesized in vitro. It is not known whether one protein species is preferentially synthesized or active in *Physarum* (Muscarella et al., 1990).

I-*PpoI* is a small homodimer of apparent molecular mass 34–39 kDa in solution (Ellison & Vogt, 1993). The endonuclease binds and induces a shallow bend in the partially symmetric, 15 bp homing site in the presence of EDTA, and can be activated by several divalent metal ions, including Mg^{2+} , Mn^{2+} , Ca^{2+} , Co^{2+} , and Zn^{2+} (Muscarella et al., 1990; Lowery et al., 1992; Ellison & Vogt, 1993; Wittmayer & Raines, 1996). Homing site binding and cleavage are salt dependent, with $K_{d/s}$ ranging from 3.3 to 112 nM in the presence of 33–275 mM NaCl. Cleavage is particularly efficient, with a k_{cat}/K_m of $10^8 M^{-1} s^{-1}$, and occurs at the center of the 15 bp homing site to generate four base, 3' overhangs of sequence 5'-TTAA-3' (Lowery et al., 1992; Wittmayer & Raines, 1996). The interaction of I-*PpoI* with its homing site has been well characterized, although the degree of sequence degeneracy tolerated within the site has not yet been defined. The enzyme appears to share some functional characteristics with a number of Type II restriction endonucleases, such as high sequence specificity and catalytic efficiency.

We have begun a structural analysis of I-*PpoI* to determine the molecular basis for homing site recognition and cleavage by I-*PpoI* and related mobile intron endonucleases. We report here the over-expression and purification of the 163 residue short form of I-*PpoI* from *E. coli*, identification of a requirement for divalent zinc ions during protein folding, and the co-crystallization and initial characterization of I-*PpoI* bound to a synthetic 21 bp homing site oligonucleotide.

Results

I-*PpoI* was subcloned, overexpressed in *E. coli*, and purified as described in Materials and methods. Initial protein expression experiments using *E. coli* grown in LB or in enriched LB media led to the production of large quantities of insoluble protein. Exhaustive attempts to purify, denature, and refold this insoluble protein were unsuccessful. We subsequently found that soluble, active I-*PpoI* endonuclease could be isolated by refolding denatured protein in the presence of 2 mM Zn^{2+} , or by expressing I-*PpoI* in *E. coli* supplemented with 2 mM Zn acetate at the time of IPTG induction. Under the latter growth conditions virtually all of the recombinant protein, representing approximately 30% of total cellular protein, partitioned into the soluble fraction. The presence of the His-Cys box motif in I-*PpoI* and the requirement for Zn to maintain protein solubility suggest that I-*PpoI* contains at least one bound zinc ion that is required for correct folding in *E. coli*.

Initial crystals of the apoenzyme that diffracted too weakly to be used for structure determination were used in a spectroscopic experiment at Brookhaven synchrotron light source, beamline X12-C,

to measure the intrinsic fluorescence of the crystals as a function of X-ray beam wavelength. The crystals are specimens of a primitive orthorhombic space group, with unit cell dimensions $a = 54 \text{ \AA}$, $b = 115 \text{ \AA}$, and $c = 240 \text{ \AA}$. The resulting spectrum (Fig. 1), indicates the presence of an absorption and fluorescence edge at 1.28 \AA (9.66 keV), corresponding to the K-edge of the zinc ion. This results indicates the presence of one or more bound zinc atoms in the purified, crystallized enzyme.

Crystals of I-*PpoI*, suitable for structure determination, were next grown complexed with a synthetic homing site oligonucleotide under conditions that support homing site binding but not cleavage. The 21-base synthetic oligonucleotide contained the 15 bp I-*PpoI* homing site and forms a palindromic, 20-base pair duplex with a 5' deoxythymidylate overhang at each end (Fig. 2). The resulting crystals were triangular and belong to space group of P3₁21. The unit cell dimensions are $a = b = 114.0 \text{ \AA}$, $c = 89.0 \text{ \AA}$. The asymmetric unit of this crystal form (which was confirmed by Patterson analysis of an initial isomorphous derivative, see below) contains one dimer of protein plus one DNA duplex with a specific volume V_m (Matthews, 1968) of $3.4 \text{ \AA}^3/\text{dalton}$. Flash-cooled native crystals diffracted to 2.5 \AA when treated as described in Materials and methods. A complete native data set has been collected to 2.7 \AA resolution (Table 1).

An initial high-quality derivative was prepared for this crystal form by co-crystallizing the enzyme with an analogous DNA construct containing a single synthetically incorporated iodine atom in each DNA strand as described in Materials and methods (Fig. 2). Difference Patterson analysis of this derivative allowed us to confirm the space group assignment and the contents of the asymmetric unit. Further derivatization trials are underway.

The hypothesis that I-*PpoI* requires a zinc atom or atoms for proper folding and/or stability was further tested by substituting 2 mM $CdCl_2$ for zinc acetate during protein expression. Although protein expression and recovery were sharply reduced, I-*PpoI* expressed and purified in the presence of Cd^{2+} was soluble and crystallized with the native homing site oligonucleotide (Fig. 2) in a space group isomorphous to that observed for the original protein preparation. Interestingly, these co-crystals were larger and diffracted to a higher resolution than co-crystals grown with endonuclease expressed and purified in the presence of Zn^{2+} . A data set has been collected from these co-crystals that is complete to 2.1 \AA

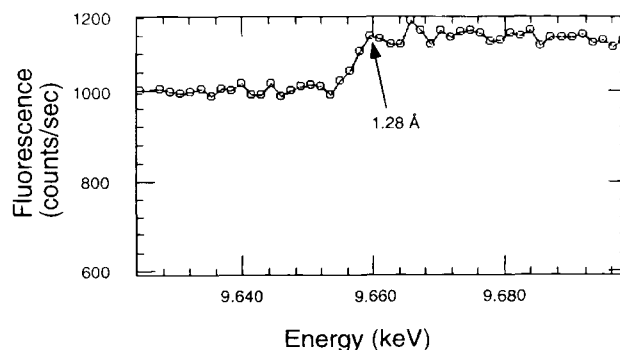


Fig. 1. Fluorescence spectrum of I-*PpoI* apo enzyme crystals. The spectrum shows a clear edge at an incident X-ray wavelength of 1.28 \AA , corresponding to an energy of 9.66 keV. The spectrum was collected at beamline X12-C at Brookhaven National Synchrotron Light Source, using a tunable monochromator.

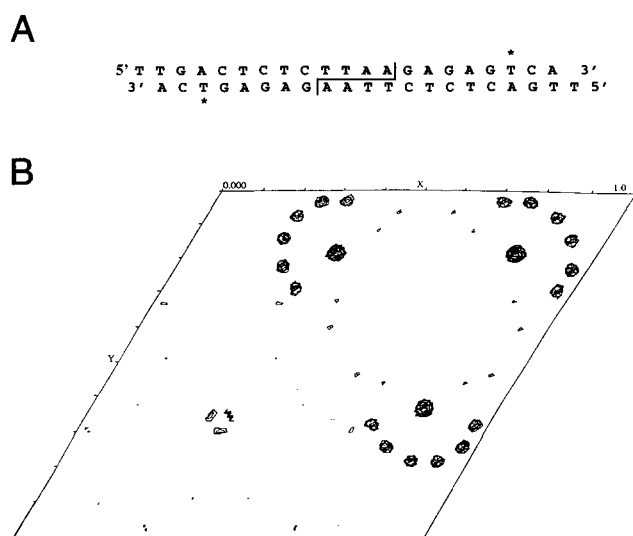


Fig. 2. **A:** Oligonucleotide used for crystallization of I-PpoI. The endonuclease and DNA were co-crystallized in the presence of EDTA, which allows binding but prevents cleavage. Asterisks indicate the positions of the deoxythymidylates that were replaced with iododeoxyuridine for isomorphous replacement. The staggered line indicates the position of the asymmetric cleavage of the homing site by I-PpoI to generate 4 base, 3' overhangs of sequence TTAA. **B:** $w = 1/3$ Harker section for isomorphous difference Patterson map of this derivative. The peaks visible are consistent with two iodine sites per asymmetric unit, one on each DNA strand. The largest peak is a doublet cross-peak. The location of the Harker peaks in this map, along with the phasing power and refined occupancy of the sites based solely on the anomalous signal, allows the space group to be unambiguously assigned as P3₁21.

resolution (Table 1). Difference Fourier and Patterson analyses indicate the presence of two pairs of distinct peaks related by a twofold non-crystallographic symmetry axis. These appear to represent the substitution of cadmium (atomic number 48) for zinc (atomic number 30). Therefore, our initial analyses appear to indicate four such sites per asymmetric unit, corresponding to two bound metals per subunit. This will be confirmed during the structure determination and refinement.

Discussion

We have overexpressed, purified, and crystallized the 18 kDa I-PpoI mobile intron endonuclease from *Physarum polycephalum* in the

presence and absence of homing site DNA. The expression of soluble protein depended critically on the presence of divalent metal ions in the growth medium. This metal ion requirement appears to reflect the requirement for one or more metal ions for proper protein folding or stability. We have confirmed the presence of zinc atoms in the crystallized apoprotein by X-ray spectroscopy, and have been able to grow co-crystals with homing site DNA using protein expressed and purified in the presence of either zinc or cadmium.

There are now several hundred known zinc-containing enzymes and proteins, many of which have been characterized in detail (Coleman, 1992; Berg & Shi, 1996). A catalytic role for zinc was first proposed for carbonic anhydrase (Keilin & Mann, 1940), and a structural role for zinc in a DNA-binding protein was first proposed and demonstrated for the transcription factor TFIIIA over 40 years later (Hanas et al., 1983; Miller et al., 1985). Small zinc-binding domains have since been identified in a wide variety of proteins involved in gene regulation; more than 10 classes of such domains have been identified (Berg & Shi, 1996). Although a complete list of functions has not been described for all of these proteins, to our knowledge no site-specific endonuclease has been described that contains such a domain. However, zinc finger motifs have been described over the past several years in a variety of related catalysts, particularly DNA and RNA polymerases (Sentenac et al., 1992; Joyce & Steitz, 1994).

The sequence of the short (163 residue) form of I-PpoI contains a region quite rich in Cys and His residues between residues 73 and 138, with the sequence H-X₄-H-X₁₅-C-X₃-H-X-C-H-X₃-C-H-X₃-H-X-C-X₁₂-C-X-C-X₆-H-X₃-C. Although this sequence has the potential to bind one or more zinc atoms per subunit, it does not align well with any consensus sequence of any previously identified zinc binding domain. Elucidation of the structure of I-PpoI should allow us to determine how Zn atoms are bound, and whether the structure of this His-Cys-rich motif is novel or is structurally related to a previously described zinc-binding domain.

Materials and methods

An I-PpoI expression vector, pET-Ppo, was constructed in three steps. The I-PpoI open reading frame was PCR amplified from plasmid p13-941 (Muscarella et al., 1990), then cleaved with the restriction endonucleases *Nde*I and *Bgl*II prior to ligation into a *Bam*HI and *Nde*I-cleaved vector, pait17. This pET11c derivative, originally constructed by W. Jack (New England Biolabs, Beverly, MA) and generously provided by D. Herrin (U. Texas-Austin), contains upstream transcription terminators and allows high level expression of the I-PpoI open reading frame under the control of T7 RNA polymerase. Protein expression was induced by IPTG induction of *E. coli* strain BL21 (DE3) (Studier et al., 1989) containing the pET-Ppo expression plasmid. Zinc acetate was added to 2 mM at the time of IPTG induction to promote the synthesis of soluble, active I-PpoI endonuclease. Cultures were induced and allowed to grow overnight at 37°C. Cells were pelleted and re-suspended in 50 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, 1 mM MgCl₂, and 5 mM DTT. The re-suspended cells can be stored at -20°C. The inclusion of Mg²⁺ at a concentration of at least 0.5 mM was essential for maintaining protein solubility through the remainder of the purification.

Recombinant I-PpoI was prepared from induced cultures by thawing and incubating cell suspensions on ice in the presence of 200 µg/mL lysozyme for at least 10 min. Cells were then lysed by

Table 1. Native data statistics

	Native	IdU	Cd
Resolution (Å)	2.7	2.7	2.1
Total reflections measured	50,053	33,285	83,203
Unique reflections	19,611	14,471	30,702
Redundancy	2.9	2.3	2.7
Completeness (%)	91.9	77.9	89.8
Average $I/\sigma(I)$	15.8	12.2	15.3
R_{merge}	0.045	0.054	0.048
Mosaicity (°)	0.30	0.25	0.37

sonication, and cell debris were removed by centrifugation. The resulting supernatant was diluted twofold with 50 mM Tris pH 8.0, then filtered through a 5.0-micron filter and loaded onto a heparin column (PerSeptive Biosystems, Cambridge, MA) equilibrated to 50 mM Tris pH 8.0, 50 mM NaCl. The protein was eluted by an increasing gradient of NaCl. The fractions containing I-PpoI were pooled and concentrated by vacuum dialysis. The concentrated protein was then passed through a Sephadex 200 column (Pharmacia, Upsala, Sweden) equilibrated with 50 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, 1 mM MgCl₂, and 5 mM DTT. I-PpoI elutes at an apparent molecular weight of approximately 35 kDa. Protein containing cadmium in the place of zinc was prepared in a similar manner with 2 mM CdCl₂ substituted for Zinc Acetate.

Crystals were grown of I-PpoI complexed with a synthetic DNA duplex (Oligos, Etc., Inc., Wilsonville, OR) that contains the 15 bp homing site. The DNA construct forms a palindromic sequence of 20 bp and contains a single 5' deoxythymidylate overhang at each end, as shown in Figure 2. The crystals were grown from a 2:1 molar ratio solution of DNA:protein in buffer from the sizing column supplemented with 2.5 mM EDTA and 5 mM spermine. The final concentration of I-PpoI in the DNA:protein complex solution was 3 mg/mL. Well solutions contained 20 mM NaCl, 2 mM EDTA, 0.1 M citrate, from pH 5.4 to 5.8, and 21% to 27% PEG4000. Equal volumes of the DNA:protein solution and the well solution were combined and suspended above the well solution using the hanging drop method of crystallization (McPherson, 1982). Optimal crystal growth was achieved at 4 °C within three to five days.

Crystals were transferred from the growth solution and transferred sequentially to 0.1 M Citrate pH 5.6, 27% PEG 4000, 20 mM NaCl, 2 mM EDTA, containing 5%, 10%, 15%, and then 20% glycerol as a cryoprotectant. A single crystal was suspended in a fiber loop, plunged into liquid nitrogen, and then maintained at 100 K during data collection. All data collection was performed on a Rigaku RAXIS II-C area detector mounted on a Rigaku RU200 rotating anode X-ray generator, $\lambda = 1.54 \text{ \AA}$ (Molecular Structure Corporation, Woodlands, TX). Data were reduced using the DENZO/SCALEPACK crystallographic data reduction package (Otwinowski, 1993). The statistics for data collection are contained in Table 1. An isomorphous derivative of native I-PpoI complexed to the synthetic oligonucleotide containing iododeoxyuridine was prepared in a manner identical to the native co-crystals, and then used to confirm the space group assignment by difference Patterson analysis.

Acknowledgments: This work was supported by NIH Grant GM49857 (B.L.S.), NIH Grant CA48022 (R.J.M.), a postdoctoral fellowship from the American Heart Association (K.E.F.), and NIH Training Grant T32 ES07032 (K.S.).

References

Belfort M, Reaban ME, Coetzee T, Dalgaard JZ. 1995. Prokaryotic introns and inteins: A panoply of form and function. *J Bacteriol* 177:3897–3903.

- Belfort M, Perlman PS. 1995. Mechanisms of intron mobility. *J Biol Chem* 270:30237–30240.
- Belfort M, Roberts R. 1997. Homing endonucleases—Keeping the house in order. *Nucleic Acids Res* 25:3379–3388.
- Berg JM, Shi Y. 1996. The galvanization of biology: A growing appreciation for the roles of zinc. *Science* 271:1081–1085.
- Coleman JE. 1992. Zinc proteins: Enzymes, storage proteins, transcription factors, and replication proteins. *Annu Rev Biochem* 61:897–946.
- Cooper AA, Stevens TH. 1995. Protein splicing: Self-splicing of genetically mobile elements at the protein level. *Trends Biochem Sci* 20:351–356.
- Duan X, Gimble F, Quioco F. 1997. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* 89:555–564.
- Ellison EL, Vogt VM. 1993. Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol Cell Biol* 13:7531–7539.
- Hanas J, Hazuda DJ, Bogenhagan DG, Wu FHY, Wu CW. 1983. *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. *J Biol Chem* 258:14120–14125.
- Heath PJ, Stephens KM, Monnat RJ, Stoddard BL. 1997. The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nat Struct Biol* 4:468–476.
- Johansen S, Embley TM, Willassen NP. 1993. A family of nuclear homing endonucleases. *Nucleic Acids Res* 21:4405.
- Johansen S, Vogt VM. 1994. An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* 76:725–734.
- Joyce CM, Steitz TA. 1994. Function and structure relationships in DNA polymerases. *Annu Rev Biochem* 63:777–822.
- Keilin D, Mann T. 1940. Carbonic anhydrase: purification and nature of the enzyme. *Biochem J* 34:1163–1176.
- Kostriken R, Strathern JN, Klar, AJ, Hicks, JB, Heffron, F. 1983. A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell* 35:167–174.
- Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. *Annu Rev Biochem* 62:587–622.
- Lowery R, Hung L, Knoche K, Bandziulis R. 1992. Properties of I-PpoI: A rare-cutting intron-encoded endonuclease. *Promega Notes* 38:8–12.
- Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol* 33:491–497.
- Miller J, McLachlan AD, Klug, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* 4:1609–1614.
- McPherson, A. 1992. *Preparation and analysis of protein crystals*. New York: John Wiley and Sons, Inc.
- Mueller JE, Brysk M, Loizos N, Belfort M. 1993. Homing endonucleases. In: Linn SM, Lloyd RS, Roberts RJ, eds. *Nucleases*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 111–143.
- Muscarella DE, Ellison EL, Ruoff BM, Vogt VM. 1990. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol Cell Biol* 10:3386–3396.
- Muscarella DE, Vogt VM. 1989. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* 56:443–454.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. In: *Methods in enzymology, Volume 276: Macromolecular crystallography, Part A*. Orlando, FL: Academic Press. pp 307–326.
- Sentenac A, Riva M, Thuriaux P, Buhler JM, Treich I, Carles C, Werner M, Ruet A, Huet J, Mann C, Chiannikulchai N, Stettler S, Mariotte S. 1992. Yeast RNA polymerase subunits and genes. In: McKnight SL, Yamamoto KR, eds. *Transcriptional regulation 1*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1989. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60–89.
- Whittmayer PK, Raines RT. 1996. Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry* 35:1076–1083.