Acta Cryst. (1998). D**54**, 1435–1436

Crystallization and preliminary crystallographic analysis of the archaeal intron-encoded endonuclease I-*Dmo*I

Jacob Z. Dalgaard,[a,b]† George H. Silva,[a,b] Marlene Belfort[a,b] and Patrick Van Roey[a,b]* *at* [a]*Wadsworth Center, New York State Department of Health, PO Box 509, Albany, NY 12201, USA, and* [b]*Department of Biomedical Sciences, School of Public Health, University at Albany, Albany, NY 12201, USA. E-mail: vanroey@wadsworth.org*

(*Received 30 January 1998; accepted 29 April 1998*)

Abstract

Two forms of the archaeal intron-encoded site-specific endonuclease I-*Dmo*I, namely I-*Dmo*Ic and I-*Dmo*Il, have been purified and crystallized. Crystals of I-*Dmo*Ic are rod-shaped and diffract to 3.0 Å resolution, but further analysis was hampered by twinning. Crystals of I-*Dmo*Il, which is a six-amino-acid C-terminal truncation of I-*Dmo*Ic, are plate shaped and belong to space group *C*2 with cell parameters $a = 93.72$, $b = 37.03$, $c = 55.56$ Å, $\beta = 113.4^\circ$, with one molecule per asymmetric unit ($V_m = 2.01$ Å^3 Da^{-1}). The crystals diffract to at least 2.3 Å resolution. A complete native data set has been measured and structure determination is on-going.

1. Introduction

I-*Dmo*I is a site-specific endonuclease encoded by an archaeal intron present in the 23S rRNA gene of the hyperthermophile *Desulfurococcus mobilis* (Kjems & Garrett, 1985; Dalgaard *et al.*, 1993). This enzyme is a member of a family of homing endonucleases characterized by the presence of a conserved repeated dodecapeptide motif named LAGLIDADG (Dalgaard *et al.*, 1997; Belfort & Roberts, 1997). I-*Dmo*I has been shown to be involved in promoting the mobility of the intron that encodes it (Aagaard *et al.*, 1995), in a similar fashion to other members of this family encoded by group I introns and inteins (Jacquier & Dujon, 1985; Gimble & Thorner, 1992). I-*Dmo*I recognizes and cleaves a 14–20 base-pair DNA sequence present in an intron-minus allele, generating a four-base-pair 3′-hydroxyl overhang (Dalgaard *et al.*, 1994; Aagaard *et al.*, 1997). The break initiates a recombinational event that utilizes the intron-containing allele as donor, leading to the propagation of the intron to the new allele.

The archaeal intron that encodes I-*Dmo*I is unusual in that it cyclizes after splicing. Since the open reading frame encoding I-*Dmo*I does not contain a stop codon 3′ to the start codon within the intron sequence, several different forms of I-*Dmo*I can theoretically be expressed *in vivo*. These include translation products of the pre-rRNA, as well as of the linear and circular introns. Two of the forms, I-*Dmo*Il (linear intron-encoded) and I-*Dmo*Ic (circular intron-encoded), which differ in length by six amino acids, have been shown to possess the same site-specific endonuclease activity (Dalgaard *et al.*, 1993) and were used in crystallization trials.

† Present address: NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, PO Box B, Building 549 Room 154, Frederick, MD 21702-1201, USA.

2. Experimental

2.1. *Purification*

Both forms of I-*Dmo*I were purified to near homogeneity after overexpression in *Escherichia coli* as previously described (Dalgaard *et al.*, 1994). Briefly, the purification was performed as follows. The enzyme was overexpressed using the pT7 expression system (Studier & Moffatt, 1986) and cells were disrupted by sonication. The enzyme was purified by ammonium sulfate fractionation, followed by chromatography using heparin-affinity and phenyl-Sepharose columns. After purification the enzyme was concentrated over a 1 ml heparin column (Hitrap Heparin, Pharmacia). Pooled fractions from the phenyl-Sepharose column were applied directly to the heparin column and eluted with a step gradient from 0.5 to 1.5 *M* NaCl. The enzyme-containing fractions were dialysed against 10 m*M* Tris–HCl pH 8.0, 1 m*M* DTT, 10% ethylene glycol and 0.5 *M* NaCl. The final concentrations were approximately 10 mg ml^{-1}. The purity and molecular weight were verified using electrospray mass spectrometry: 22609 Da for I-*Dmo*Ic and 22003 Da for I-*Dmo*Il. The secondary structure composition of I-*Dmo*Ic was estimated from a circular dichroism spectrum. The spectrum was obtained using a J-720 Spectropolarimeter (JASCO) that had been calibrated with aminocamphorsulfonate using a 0.5 mm path length. The sample contained 0.5 mg ml^{-1} I-*Dmo*Ic in 20 m*M* Tris–HCl and 100 m*M* NH_4OAc. Using the method of Yang *et al.* (1986), the composition was estimated to be 50.2% β-sheet, 28.1% α-helix, 21.6% random coil and 0% β-turn.

2.2. *Crystallization*

Crystals were grown by hanging-drop vapor-diffusion methods. An initial screening for crystallization conditions was performed with the sparse-matrix sampling protocol (Jancarik & Kim, 1991), using drops containing 2 µl protein solution and 2 µl precipitating solution equilibrated against 800 µl of reservoir solution at 293 K. For I-*Dmo*Ic, small needles formed after one day under several conditions of the Crystal Screen I (Hampton Research) kit, all of which contained PEG as the primary precipitating agent. After three to four weeks, several high-salt solutions also yielded needle-shaped crystals. Optimization experiments led to the following conditions for crystallization: drops containing 3 µl protein at 10 mg ml^{-1} in 5 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 10% ethylene glycol, 1 m*M* DTT and 3 µl precipitating buffer were equilibrated against 1 ml of precipitating buffer containing 2–3% PEG 3350, 1–1.2 *M* sodium acetate, 10% ethylene glycol, 1% 2-methyl-2,4-pentanediol, 50 m*M* Na^+, K^+ phosphate pH 8.0 and 100 m*M* Tris–HCl pH 8.0. The largest crystals, up to 1.0 × 0.1 × 0.1 mm, were obtained by initiating crystallization

© 1998 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0907-4449 © 1998

experiments at 310 K and by stepwise lowering of the temperature to 293 K over a period of four weeks. The crystals are rod-shaped and diffract to at least 3.0 Å resolution, but further analysis was hampered by twinning problems.

For I-*Dmo*II, initial crystallization trials with the Crystal Screen kit yielded small crystals from 30% PEG 3350, 200 m*M* Li_2SO_4 in 100 m*M* Tris–HCl pH 8.5 using protein at approximately 10 mg ml^{-1} in 5 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 10% ethylene glycol and 1 m*M* DTT. Optimization experiments resulted in reproducible large crystals from drops containing 4 µl protein solution as described above, 4 µl precipitating buffer and 1 µl of 20 m*M* *n*-octyl-β-glucopyranoside equilibrated against 1 ml of a precipitating buffer consisting of 30% PEG 3350, 200 m*M* sodium acetate in 100 m*M* Tris–HCl pH 8.5. Plate-shaped crystals, as shown in Fig. 1, grow in 2–5 days to dimensions of 0.5 × 0.3 × 0.1 mm.

3. Results and discussion

X-ray diffraction data were measured on a Rigaku R-AXIS IIc detector equipped with mirror optics and a liquid-nitrogen low-temperature device. The crystal was flash-cooled to approximately 130 K. No additional cryoprotectant was required. The I-*Dmo*II crystals belong to the monoclinic space group *C*2 with cell parameters a = 93.72, b = 37.03, c = 55.56 Å, β = 113.4° and V = 176 960 Å^3. With a molecular weight of 22 003 Da and one molecule per asymmetric unit, V_m is 2.01 Å^3 Da^{-1}, corresponding to a solvent content of 33.3% (Matthews, 1968). A 2.33 Å native data set was measured using a crystal with approximate dimensions 0.4 × 0.3 × 0.1 mm. The data were processed and scaled with the *BIOTEX* package (Molecular Structure Corporation). Details of the data-reduction statistics are listed in Table 1.

The structures of two other LAGLIDADG proteins, the homing endonuclease I-*Cre*I (Heath *et al.*, 1997) and the intein PI-*Sce*I (Duan *et al.*, 1997), were recently published. I-*Cre*I is unusual in that it contains only one LAGLIDADG motif and functions as a homodimer. The structure of I-*Cre*I shows the two monomers related by a crystallographic twofold rotation axis. PI-*Sce*I consists of a protein-splicing domain and an endonuclease domain. The endonuclease domain contains two LADLIDADG motifs and has a two-subdomain fold. The overall topology of each subdomain resembles that of an I-*Cre*I monomer but the subdomains differ significantly in the dimensions and relative orientations of some of the secondary elements, resulting in an asymmetric structure. I-*Dmo*I differs from both these enzymes since it contains two LAGLIDADG motifs but no protein-splicing domain. A self-rotation search, using the program *AMoRe* (Navaza, 1994), was performed for I-*Dmo*I. Since no twofold rotation peak other than the crystallographic rotation axis could be detected, the overall structure of I-*Dmo*I is likely to be more similar to that of the endonuclease domain of PI-*Sce*I than to that of the I-*Cre*I dimer. Molecular-replacement calculations aimed at determining the structure of I-*Dmo*I, using the structure of I-*Cre*I as a model, have not yielded any valid solutions. This may be explained by the fact that the I-*Cre*I dimer is much larger than I-*Dmo*I (276 compared with 187 residues) but is probably also a consequence of the anticipated lower symmetry of the latter. Therefore, the structure of I-*Dmo*I will be determined by multiple isomorphous replacement or multiple anomalous diffraction (MAD) methods. Currently, at least two potential heavy-atom derivatives have been identified; however, phasing attempts are hampered by non-isomorphism problems. Therefore, MAD phasing is the most likely to be successful.

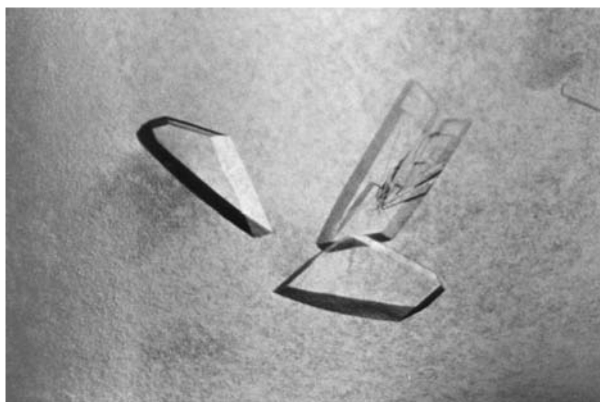

Fig. 1. Plate-shaped crystals of I-*Dmo*II. The crystals shown have approximate dimensions of 0.4 × 0.25 × 0.1 mm.

Table 1. *Data reduction statistics for the 233 Å resolution data set for I-DmoII*

	All data, 20–2.33 Å	High-resolution shell, 2.5–2.33 Å
Total observations	30357	5409
Unique observations	7800	1464
Completeness (%)	95.2	93.3
R_{merge}	0.061	0.188
$\langle F^2/\sigma(F^2)\rangle$	13.52	3.84
$F^2/\sigma(F^2) > 2$ (%)	82.1	59.9

Research supported in part by NIH grants GM39422 and GM44844 (MB) and GM56966 (PVR). JZD was supported by the Danish Natural Science Research Council. Circular dichroism and mass spectrometry measurements were performed in the Wadsworth Center's Biochemistry and Mass Spectrometry Core Facilities.

References

Aagaard, C., Awayez, M. J. & Garrett, R. A. (1997). *Nucleic Acids Res.* **25**, 1523–1530.
Aagaard, C., Dalgaard, J. Z. & Garrett, R. A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 12285–12289.
Belfort, M. & Roberts, R. J. (1997). *Nucleic Acids Res.* **25**, 3379–3388.
Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 5414–5417.
Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1994). *J. Biol. Chem.* **269**, 28885–28892.
Dalgaard, J. Z., Klar, A., Moser, M., Holley, W. R., Chatterjee, A. & Mian, S. (1997). *Nucleic Acids Res.* **25**, 4626–4638.
Duan, X., Gimble, F. S. & Quiocho, F. A. (1997). *Cell*, **89**, 555–564.
Gimble, F. S. & Thorner, J. (1992). *Nature (London)*, **357**, 301–306.
Heath, P. J., Stephens, K. M., Monnat, R. J. Jr & Stoddard, B. L. (1997). *Nature Struct. Biol.* **4**, 468–476.
Jacquier, A. & Dujon, B. (1985). *Cell*, **41**, 383–394.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 490–411.
Kjems, J. & Garrett, R. A. (1985). *Nature (London)*, **318**, 675–677.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.
Yang, J. T., Wu, C. & Martinez, H. M. (1986). *Methods Enzymol.* **130**, 208–269.