

DNA homing site recognition and cleavage by I-PpoI, a nuclear intron-encoded homing endonuclease

Karen E. Flick¹, Raymond J. Monnat, Jr.², and Barry L. Stoddard¹

¹ Fred Hutchinson Cancer Research Center A3-023 1100 Fairview Ave. N
Seattle, WA 98019, USA

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

INTRODUCTION

Homing endonucleases are a diverse family of proteins encoded by open reading frames in mobile self-splicing introns. Similar homing endonucleases have also been identified as optional, independently folded domains in self-splicing protein introns, termed 'inteins', and as free-standing open reading frames (ORFs). Homing endonucleases promote the lateral transfer of their encoding introns or open reading frames by a targeted transposition mechanism termed intron mobility or homing. This process is initiated and targeted by the endonuclease, which makes a highly site-specific DNA double strand break at a homing site that lacks the self-splicing intron or open reading frame (an "intron- or ORF-less" allele). The endonuclease ORF is subsequently transferred to the cleaved recipient allele by a double strand break repair/gene conversion event (Figure 1). Homing endonucleases have been identified in a diverse collection of unicellular eukaryotes, *Archea*, and eubacteria, with the largest number having been identified in organellar genomes. These comparatively small endonucleases (18 to 22 kDa per subunit for active dimers, 32 to 44 kDa for active monomers) share the ability to recognize and cleave long DNA target or "homing" sites of 15 to 40 bp with high specificity *in vivo* and *in vitro* [1-5].

Four homing endonuclease families have been identified on the basis of conserved protein motifs. These include the LAGLIDADG, His-Cys box, GIY-YIG and HNH endonucleases. The His-Cys box endonucleases contain two conserved histidines and three conserved cysteines within a 30 residue region of protein (Figure 1). The first His-Cys box homing endonuclease identified, as well as the first nuclear homing endonuclease, was I-PpoI [6-8]. The intron encoding I-PpoI is one of three mobile introns known to reside in a eukaryotic nucleus. The other two nuclear-encoded endonucleases were identified in nuclear introns of the slime mold *Didymium iridis*, and in the amoeba *Naegleria andersoni* ssp *andersoni* [9]. The His-Cys box endonucleases, like the LAGLIDADG homing endonucleases, cleave their homing site DNAs to generate 4 base, 3' extended cohesive ends. The remaining two families of homing endonucleases consist of the bacteriophage-encoded GIY-YIG endonucleases that cleave homing site DNA to generate 2 base, 5' extended cohesive ends; and the catalytically diverse H-N-H homing endonucleases, that cleave their homing site DNAs to generate 2 base or 10 base, 5' extended cohesive ends, or that cleave only one of the strands of homing site DNA.

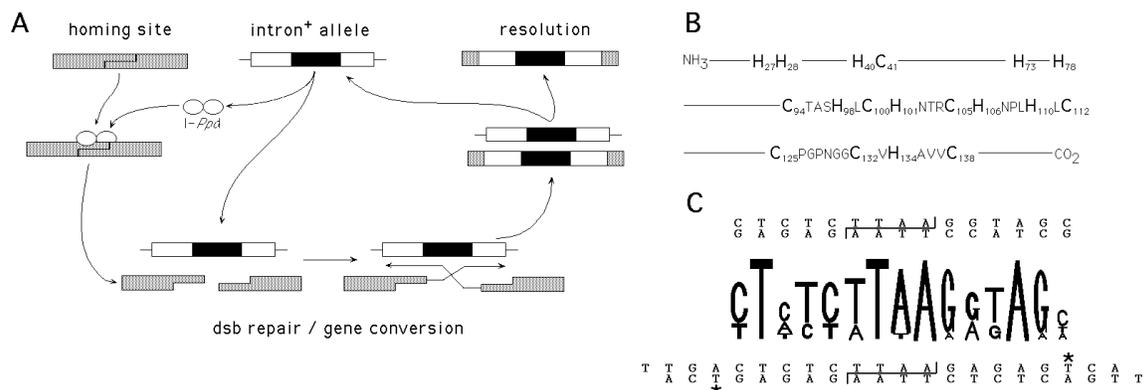


Figure 1. **The *I-PpoI* intron homing cycle and sequences of the *I-PpoI* homing site and His-Cys box motif.** A: Intron mobility or homing is initiated and targeted by a site-specific DNA double strand break made in an intron-less allele B: Diagrammatic representation of the primary structure of the *I-PpoI* protein. The conserved His-Cys boxes are found between residues 73 and 138. C: Sequence of the wild-type, pseudo-palindromic *I-PpoI* homing site (above), and the synthetic DNA palindrome used for cocrystallization (below). The position of the thymine nucleotides derivatized to iodinated uracils is shown by asterisks; the cleavage pattern (which produces 4 base, 3' extended ends) is indicated by the staggered underline. Aligned in between the wild type and synthetic homing site sequences is a code denoting relative invariance of positions in the *I-PpoI* homing site.

We have expressed, purified and crystallized recombinant *I-PpoI* homing endonuclease with a 20 base pair synthetic *I-PpoI* DNA homing site (Figure 1) [10]. Preliminary X-ray studies indicated that the protein contains two bound zinc ions per subunit that are essential for protein folding. A requirement for zinc to allow *I-PpoI* folding or activity was suggested by the presence of the potential zinc binding His-Cys protein motif.

METHODS AND THE USE OF ALS FACILITIES

Crystals were grown of *I-PpoI* complexed with a synthetic DNA duplex (Oligos, Etc., Inc.) that contains a variant of the 15 bp *I-PpoI* homing site (Figure 1). The DNA construct forms a palindromic sequence of 20 bp and contains a single 5' deoxythymidylate overhang at each end. A high resolution native data set, used to refine the *I-PpoI* structure to its current resolution value of 1.8 Å, was collected on an ADSC 4 panel CCD area detector at beamline 5.0.2 at the Advanced Light Source (LBNL, Berkeley, CA) with the assistance and advice of Thomas Earnest, using a wavelength of 1.0 Å. This data set, which extends to 1.6 Å resolution, represents a dramatic improvement from a previous best of 2.4 Å, which was collected earlier on a rotating anode generator equipped with mirrors. All RAXIS data were reduced using the DENZO/SCALEPACK crystallographic data reduction package [11], while the ALS CCD data were reduced using MOSFLM 5.5 [12]. Two isomorphous derivatives were prepared for structure determination. The first was prepared by synthetic incorporation of an iododeoxyuridine in the DNA duplex (Figure 1). A second derivative was prepared by a 1 hour soak of the native crystals in a 1 mM $\text{Hg}(\text{CN})_2$ solution prior to flash-cooling. All initial MIR phase calculations and phase refinements were done using the CCP4 package [13], with phases calculated using the program MLPHARE. Phases calculated using these data yielded an interpretable electron density map that was further improved by solvent flattening, histogram matching, and phase extension. Phase-combined maps were calculated using SFALL and SIGMAA to resolve ambiguous regions of the original DM electron density map. The resolution used for the refinement of the final models extends to 1.8 Å. There is a single *I-PpoI* dimer and DNA duplex in the asymmetric unit of the crystal.

RESULTS

I-PpoI monomer folds into a single extended domain, of dimensions 25 Å x 35 Å x 55 Å, that contains three separate, anti-parallel β-sheets flanked by two long α-helices. The extended fold of the enzyme monomer is stabilized by two covalently bound, independently coordinated zinc ions located 15 Å apart and on the protein face opposite the DNA binding surface. The homodimer contains four bound zinc ions.

The *I-PpoI* endonuclease dimer forms an extended DNA binding surface that is approximately 80 Å long and oriented along the DNA-binding axis (Figure 2). The homodimer is stabilized by an unusual combination of a very small, highly solvated central interface, and an extended C-terminal tail that extends 34 Å across the adjacent monomer. *I-PpoI*, in contrast to the LAGLIDADG homing endonucleases, does not exhibit a tightly packed dimer interface that facilitates homing site cleavage across the minor groove.

The two bound zinc ions in each protein monomer are tetrahedrally coordinated in different motifs composed of cysteine and histidine residues. The two motifs are distinct from one another, and are also differ from previously described zinc binding motifs. The first zinc ion is coordinated by a Cys₃-His₁ ligand cluster from the N-terminal end of the protein, with a single side chain contributed by β2 (Cys 41) and the remaining three side chains by a short loop between β7 and β8 (Cys 100, Cys 105, and His 110). The binding motif for this metal is therefore C-X₅₈-C-X₄-C-X₄-H. The second zinc ion is coordinated by a cluster of four closely spaced protein side chains (Cys₂-His₁-Cys₁ coordination), all donated from a short buried protein loop (Cys 125, Cys 132, His 134, and Cys 138) near the C-terminal end of the enzyme. The motif for the second bound zinc ion is C-X₆-C-X₁-H-X₃-H.

The 15 bp *I-PpoI* homing site is one of the shortest DNA homing sites characterized to date. This is nonetheless a large target for sequence-specific recognition and binding by the 163 residue *I-PpoI* protein. An important part of the DNA recognition strategy of *I-PpoI* is to use a homodimer as the active form of the endonuclease to recognize the limited two-fold sequence symmetry within the homing site. The primary contacts made by *I-PpoI* to homing site DNA are between the second β-sheet (β3-β4-β5) of each monomer and base pairs in the major groove to either side of the cleavage site at positions -5 to -9 and +5 to +9. Additional contacts are made in the center of the complex, across the minor groove, and between residues in the active site and DNA atoms in the cleavage site. The *I-PpoI* homodimer also strongly bends homing site DNA to facilitate cleavage across the minor groove with the generation of 4 base, 3' extended cohesive ends.

The structure determined with ALS data clearly indicates a divalent cation is bound in the active site and the scissile phosphate has been cleaved. The metal ion is coordinated in both structures to the 3' hydroxyl leaving group of the phosphate backbone and the side chain oxygen of Asn 119, and has four easily visible coordinating water molecules, one of which directly contacts an oxygen atom of the cleaved 5' phosphate group. The distance between the 3' hydroxyl oxygen leaving group and the 5' phosphorus atom refines to 3.5 to 4.3 Å in all of the structures. Apart from the active site divalent cation and its coordinating asparagine residue, the other groups within potential chemical interaction distance of the scissile phosphate group are His 98 (which is absolutely conserved in the SHLC sequence found in the His-Cys box endonucleases), Arg 61 (which contacts oxygen atoms on the cleaved phosphate and its immediate phosphate neighbor through its two guanidinyll nitrogens), and His 78.

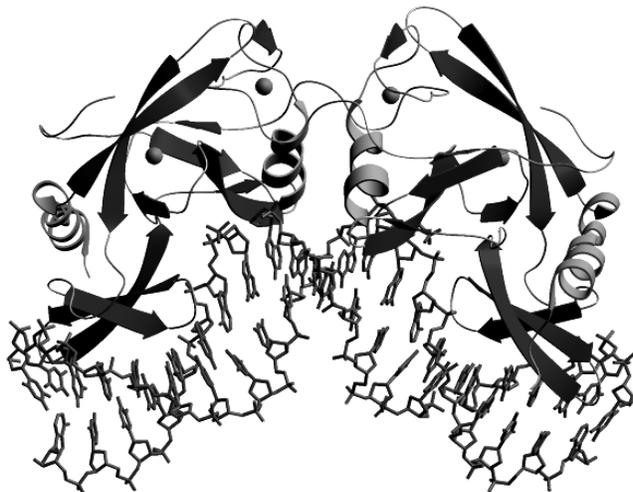


Figure 2. Stereo ribbon diagram of the endonuclease dimer in complex with the DNA homing site. The substrate is dramatically deformed in the complex, with a sharp pair of kinks near the cleavage site opening the major groove, and a pair of shallower opposing bends to either side giving an overall bend of approximately 50 degrees. This figure made with program SETOR [14].

REFERENCES

1. M. Belfort and R. Roberts, *Nucleic Acids Res.* **25** (17), 3379 - 3388 (1997).
2. A. M. Lambowitz and M. Belfort *Ann. Rev. Biochem.* **62**, 587 - 622 (1993).
3. J. E. Mueller, M. Brysk, N. Loizos and M. Belfort, in *Nucleases*, 2nd Ed., edited by S. M. Linn, R. S. Lloyd, and R. J. Roberts (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 111-143 (1993).
4. M. Belfort and P. S. Perlman, *J. Biol. Chem.* **270**, 30237 - 30240 (1995).
5. M. Belfort, M. E. Reaban, T. Coetzee, and J. Z. Dalgaard *J. Bacter.* **177**, 3897 - 3903 (1995).
6. D. E. Muscarella and V. M. Cell **56**, 443 - 454 (1989).
7. D. E. Muscarella, E. L. Ellison, B. M. Ruoff, and V. M. Vogt, *Mol. and Cell. Biol.* **10**(7), 3386 - 3396 (1990).
8. S. Johansen and V. M. Vogt, *Cell* **76**(4), 725 - 734 (1994).
9. S. Johansen, T. M. Embley and N. P. Willassen, *Nuc. Acids Res.* **21** (18), 4405 (1993).
10. K. E. Flick, D. McHugh, J. D. Heath, K. M. Stephens, R. J. Monnat, Jr. and B. L. Stoddard, *Protein Science* **6**, 1 - 4 (1997).
11. Z. Otwinowski and W. Minor, *Methods in Enzymology* **276**, 307-326 (1997).
12. A. G. W. Leslie in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography No. 26* (Daresbury Laboratory, Warrington, U.K.) (1992).
13. CCP4 The SERC (UK) Collaborative Computing Project No. 4, a suite of programs for protein crystallography. (Daresbury Laboratory, Warrington WA4 4AD, UK) (1979).
14. Evans, S. V. *J. Mol. Graphics* **11**, 134-138 (1993).

This work was supported by the National Institutes of Health under grant No. GM49857.

Principal investigator: Barry L. Stoddard, Fred Hutchinson Cancer Research Center. Email: bstoddar@fred.fhrc.org. Telephone: 206-667-4031.