0.5 M NaOH. The cell extracts were neutralized with HCl and ^{3}H radioactivity was measured in a liquid scintillation β-counter (Aloka).

Western blot analysis, immunoprecipitation, metabolic labelling, and pull-down assays. Cells indicated were lysed in TNE buffer[26]. Whole-cell extracts or immunoprecipitates produced with anti-Flag, anti-haemagglutinin (HA) or anti-Evi-1 were visualized by immunoblotting with anti-Evi-1 (ref. 7), anti-Rb (Pharmingen), anti-Flag (M2, Eastman Kodak) or anti-HA (12CA5, Boehringer) antibodies. For ligand-induced association assays, COS7 cells transiently transfected with the indicated constructs were washed and incubated in the presence of 200 pM TGF-β for 2 h. [^{32}P]phosphate labelling was performed as described[26]. For pull-down assays, 5 μg of GST fusion proteins were collected on glutathione–Sepharose beads (Pharmacia) and incubated with [^{35}S]methionine-labelled Smad3 that was synthesized using TNT-coupled reticulocyte lysate systems (Promega).

Electrophoretic mobility shift assays. EMSAs were performed as described[29], except that the oligonucleotide ATGAGTCAGACACCTCTGGCTGTCTGGA-AGGGATGAGTCAGACACCTCTG GCTTTCTGGA was used as a probe. For antibody supershift assays, 3 μg anti-Flag (M2) antibody was added to the reaction mixture.

Transcription-response assays. Transcriptional-response assays were performed as described[7], except that we used LipofectAmine (Gibco BRL) instead of a calcium phosphate method. As an internal control of transfection efficiency, a plasmid expressing β-galactosidase was co-transfected. Cells were treated for 24 h with TGF-β; relative luciferase activity was then measured in cell extracts using the luciferase assay system (Promega) and a luminometer (Lumat, Berthold) and was normalized to the β-galactosidase activity.

Immunofluorescence. Cells were treated with anti-Flag and anti-Evi-1 antibodies and then incubated with fluorescein isothiocyanate(FITC)-conjugated donkey anti-mouse IgG and Texas-red-conjugated donkey anti-rabbit IgG (Jackson Immunologicals), respectively, as secondary antibodies. Images were obtained using a confocal mciroscope (MRC-1024, Bio-Rad). The cells were counterstained with Hoechst 33258 to identify nuclei and evaluate transfection efficiency.

Received 25 March; accepted 27 April 1998.

1. Lopingco, M. C. & Perkins, A. S. Molecular analysis of Evi1, a zinc finger oncogene involved in myeloid leukemia. *Cur. Top. Microbiol. Immunol.* **211,** 211–222 (1996).
2. Morishita, K. *et al.* Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300–400 kilobases on chromosome band 3q26. *Proc. Natl Acad. Sci. USA* **89,** 3937–3941 (1992).
3. Mitani, K. *et al.* Generation of the AML1-EVI-1 fusion gene in the t(3;21)(q26;q22) causes blastic crisis in chronic myelocytic leukemia. *EMBO J.* **13,** 504–510 (1994).
4. Ogawa, S. *et al.* Structurally altered Evi-1 protein generated in the 3q21q26 syndrome. *Oncogene* **13,** 183–191 (1996.
5. Ogawa, S. *et al.* Increased Evi-1 expression is frequently observed in blastic crisis of chronic myelocytic leukemia. *Leukemia* **10,** 788–794 (1996).
6. Morishita, K. *et al.* Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* **54,** 831–840 (1988).
7. Tanaka, T. *et al.* Evi-1 raises AP-1 activity and stimulates c-fos promoter transactivation with dependence on the second zinc finger domain. *J. Biol. Chem.* **269,** 24020–24026 (1994).
8. Bartholomew, C., Kilbey, A., Clark, A. M. & Walker, M. The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation. *Oncogene* **14,** 569–577 (1997).
9. Kurokawa, M. *et al.* The AML1/Evi-1 fusion protein in the t(3;21) translocation exhibits transforming activity on Rat1 fibroblasts with dependence on the Evi-1 sequence. *Oncogene* **11,** 833–840 (1995).
10. Morishita, K., Parganas, E., Matsugi, T. & Ihle, J. N. Expression of the Evi-1 zinc finger gene in 32Dc13 myeloid cells blocks granulocytic differentiation in response to granulocyte colony-stimulating factor. *Mol. Cell. Biol.* **12,** 183–189 (1992).
11. Kreider, B. L., Orkin, S. H. & Ihle, J. N. Loss of erythropoietin responsiveness in erythroid progenitors due to expression of the Evi-1 myeloid-transforming gene. *Proc. Natl Acad. Sci. USA* **90,** 6454–6458 (1993).
12. Massagué, J. The transforming growth factor-β family. *Annu. Rev. Cell Biol.* **6,** 597–641 (1990).
13. Zhang, Y., Feng, X., We, R. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF-β response. *Nature* **383,** 168–172 (1996).
14. Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF-β receptor. *Nature* **370,** 341–347 (1994).
15. Macias-Silva, M. *et al.* MADR2 is a substrate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell* **87,** 1215–1224 (1996).
16. Hannon, G. J. & Beach, D. p15^{INK4B} is a potential effector of TGF-β-induced cell cycle arrest. *Nature* **371,** 257–261 (1994).
17. Laiho, M., DeCaprio, J. A., Ludlow, J. W., Livingston, D. M. & Massagué, J. Growth inhibition by TGF-β linked to suppression of retinoblastoma protein phosphorylation. *Cell* **62,** 175–185 (1990).
18. Derynck, R. & Zhang, Y. Intracellular signalling: the mad way to do it. *Curr. Biol.* **6,** 1226–1229 (1996).
19. Massagué, J. TGFβ signaling: receptors, transducers, and Mad proteins. *Cell* **85,** 947–950 (1996).
20. Wrana, J. L. & Attisano, L. MAD-related proteins in TGFβ signaling. *Trends Genet.* **12,** 493–496 (1996).
21. Wu, R.-Y., Zhang, Y., Feng, X.-H. & Derynck, R. Heteromeric and homomeric interactions correlate with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. *Mol. Cell. Biol.* **17,** 2521–2528 (1997).
22. Nakao, A. *et al.* TGF-β receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J.* **16,** 5353–5362 (1997).
23. Wieser, R., Wrana, J. L. & Massagué, J. GS domain mutations that constitutively activate T β R-I, the downstream signaling component in the TGF-β receptor complex. *EMBO J.* **14,** 2199–2208 (1995).
24. Yingling, J. M. *et al.* Tumor suppressor Smad4 is a transforming growth factor β-inducible DNA binding protein. *Mol. Cell. Biol.* **17,** 7019–7028 (1997).
25. Tanaka, T. *et al.* Dual functions of the AML1/Evi-1 chimeric protein in the mechanism of leukemogenesis in t(3;21) leukemias. *Mol. Cell. Biol.* **15,** 2383–2392 (1995).
26. Lagna, G., Hata, A., Hemmati, B. A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF-β signalling pathways. *Nature* **383,** 832–836 (1996).
27. Hata, A., S., L. R., Wotton, D., Lagna, G. & Massagué, J. Mutations increasing autoinhibition inactive tumor suppressors Smad2 and Smad4. *Nature* **388,** 82–87 (1997).
28. Takebe, Y. *et al.* SRα promoter: an efficient and versatile mammalian cDNA expression system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. *Mol. Cell. Biol.* **8,** 466–472 (1988).
29. Kurokawa, M. *et al.* A conserved cysteine residue in the *runt* homology domain of AML1 is required for the DNA binding ability and the transforming activity on fibroblasts. *J. Biol. Chem.* **271,** 16870–16876 (1996).

Acknowledgements. We thank J. Massagué for Smad1–Flag and Smad4–HA; J. L. Wrana for Smad2–Flag and the pCMV5 vector; R. Derynck for Smad3–Flag and Smad4-Flag; X.-F. Wang for p15P113–Luc; K. Miyazono for p3TP–Lux and TβRI TD; and K. Arai for pME18Sneo. This work was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Health and Welfare and from the Ministry of Education, Science, and Culture of Japan.

Correspondence and requests for materials should be addressed to H.H. (e-mail: hhirai-tky@umin.u-tokyo.ac.jp).

DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I

Karen E. Flick*, Melissa S. Jurica*†, Raymond J. Monnat Jr‡ & Barry L. Stoddard*

* *Fred Hutchinson Cancer Research Center and † Molecular and Cellular Biology Program of the University of Washington, 1100 Fairview Avenue North, Seattle, Washington 98109, USA*

‡ *University of Washington, Department of Pathology, Box 357705, Seattle, Washington 98195, USA*

Homing endonucleases are a diverse collection of proteins that are encoded by genes with mobile, self-splicing introns[1–3]. They have also been identified in self-splicing inteins (protein introns)[4]. These enzymes promote the movement of the DNA sequences that encode them from one chromosome location to another; they do this by making a site-specific double-strand break at a target site in an allele that lacks the corresponding mobile intron[3]. The target sites recognized by these small endonucleases are generally long (14–44 base pairs). Four families of homing endonucleases have been identified, including the LAGLIDADG, the His–Cys box, the GIY–YIG and the H–N–H endonucleases[1]. The first identified His–Cys box homing endonuclease was I-*Ppo*I from the slime mould *Physarum polycephalum*[5,6]. Its gene resides in one of only a few nuclear introns known to exhibit genetic mobility[7]. Here we report the structure of the I-*Ppo*I homing endonuclease bound to homing-site DNA determined to 1.8 Å resolution. I-*Ppo*I displays an elongated fold of dimensions 25 × 35 × 80 Å, with mixed α/β topology. Each I-*Ppo*I monomer contains three anti-parallel β-sheets flanked by two long α-helices and a long carboxy-terminal tail, and is stabilized by two bound zinc ions 15 Å apart. The enzyme possesses a new zinc-bound fold and endonuclease active site. The structure has been determined in both uncleaved substrate and cleaved product complexes.

The structure of I-*Ppo*I (Fig. 1) shows that the central interface of the enzyme dimer is small and highly solvated, with subunit contacts that bury only approximately 700 Å^2 of surface area (Fig. 2a). The C-terminal tails (residues 146–163) are domain-swapped and extend 34 Å across opposite monomers, burying an additional 900 Å^2 per subunit (Fig. 2b). In contrast to the LAGLIDADG homing endonucleases[8,9], the I-*Ppo*I dimer interface is not tightly packed.

The first bound zinc ion (Fig. 3) is coordinated by a cluster of three Cys and one His ligands. One side chain (Cys 41) is contributed by an amino-terminal β-strand (β2) and the remaining three side chains (Cys 10, Cys 105 and His 110) are from a short loop between β7 and β8. The binding for this metal is C-X_{58}-C-X_4-C-X_4-H. The second zinc ion is coordinated by a short cluster of four side chains (Cys 125, Cys 132, His 134 and Cys 138). The motif for the second bound zinc ion is C-X_6-C-X_1-H-X_3-H, and resembles the zinc-binding sequence from retroviral nucleocapsid proteins[10] but has a different structural fold. All eight residues involved in zinc

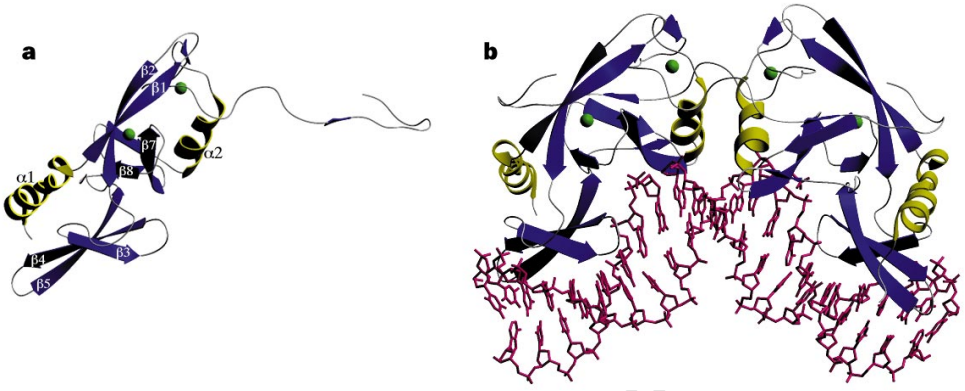

Figure 1 Tertiary fold and DNA complex of the I-*Ppo*I endonuclease. **a**, The protein monomer is stabilized by two bound zinc ions and has a long C-terminal tail that extends around the surface of the second subunit in the dimer. **b**, The bound endonuclease dimer dramatically deforms the homing-site DNA, inducing a sharp pair of kinks near the cleavage site opening the minor groove. The overall bend is approximately 50°. Figs 1, 2, 3 and 5 were made using the program SETOR[27].

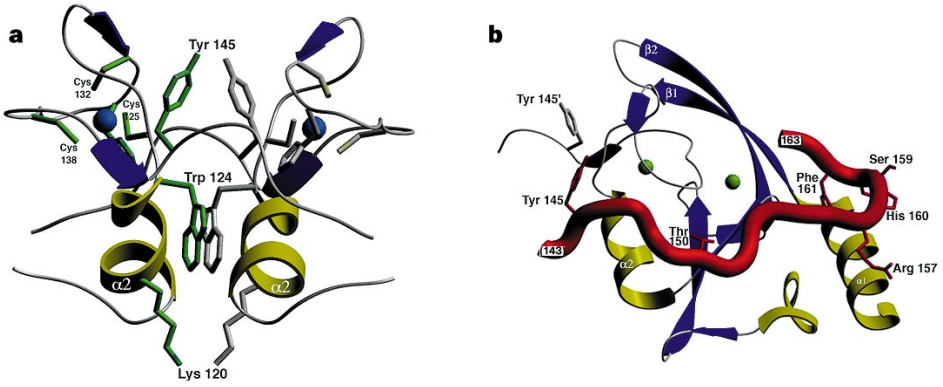

Figure 2 Structure of the dimer interface. **a**, The central dimer interface is composed primarily of Van der Waals contacts between Trp 124 and 124′. This interface buries a total of less than 700 Å^2 of protein surface. Lys 120 and 120′ make direct contacts with thymine bases in the minor groove of the central cleavage site. **b**, The domain-swapped C-terminal tail (red) crosses the dimer interface at Gly 146 and proceeds across the surface of the second subunit. Each arm results in an additional 900 Å^2 of buried surface area. Residues 150–153 form a short β-strand interaction with residues 111′–113′. Each of the C-terminal oxygens forms a hydrogen bond with His 40. Phe 160 is also buried in this interface.

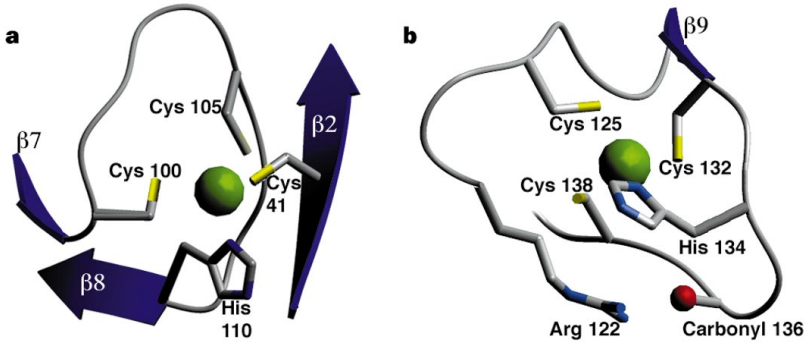

Figure 3 Zinc-binding motifs of I-*Ppo*I. **a**, The N-terminal bound zinc is coordinated by Cys 41 from a long β-sheet and three additional residues in a short loop. **b**, The C-terminal zinc is coordinated by three cysteine and one histidine residues contained within a short 14-residue sequence. This loop is flanked by a hydrogen bond between Arg 122 and the carboxyl oxygen of Val 136.

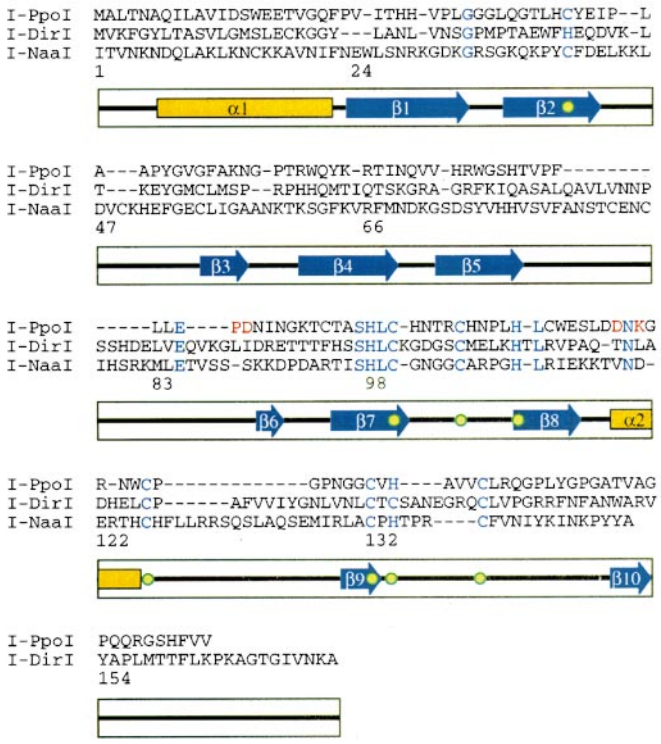

Figure 4 Sequence alignment of I-*Ppo*I and other known nuclear His–Cys box homing endonucleases. The I-*Ppo*I protein sequence is shown aligned with sequences from the homologues I-*Dir*I and I-*Naa*I and its secondary structure elements. Residues that act as ligands to the two zinc atoms are denoted by circles. Conserved residues are blue; the residues in the previously noted 'endonuclease' PD...(E/D)X(K/R) motif in the I-*Ppo*I sequence are red.

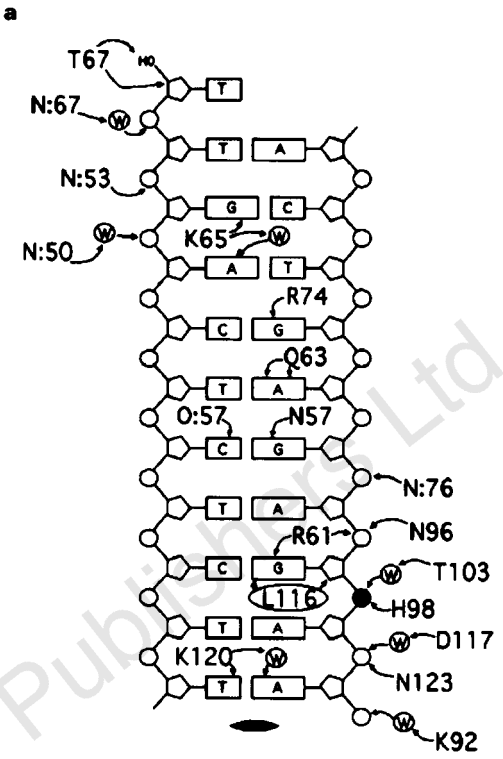

Figure 5 The endonuclease DNA-binding surface. **a**, Schematic diagram of contacts between an endonuclease monomer and a half-homing site. Residues 61, 63, 65 and 67 are all from the same face of a single β-strand and form a series of contacts down the major groove. **b**, The DNA-binding interactions of the I-*Ppo*I β-ribbon in a homing half-site major groove. The alternating side chains from β4 are shown.

coordination are conserved as Cys or His residues among the three known His–Cys box homing endonucleases[7] (Fig. 4). The conserved sequence Ser 97–His 98–Leu 99–Cys 100 contains one zinc-coordinating cysteine and an active-site histidine.

Previous structural studies have visualized three DNA-binding protein families containing two bound zinc ions per protein subunit: steroid receptors[11], the GAL4 transcription factor zinc-binding domain[12], and the RING finger motif[13–15]. Of these structures, only the steroid receptors coordinate two zinc atoms without sharing ligands between metals or using overlapping metal-binding sequences, as seen for I-*Ppo*I[11]. Rather than being directly involved in DNA binding, the I-*Ppo*I zinc-binding motifs are critical primarily for structural stabilization of the protein core.

The primary sequence-specific contacts made by I-*Ppo*I to homing-site DNA are between residues in the second β-sheet (β3–β4–β5) of each enzyme monomer and base pairs 5–9 in the major groove of each half-site. Additional contacts are made in the centre of the complex within the minor groove and with several phosphate groups in the cleavage site. The I-*Ppo*I homodimer strongly bends homing-site DNA for easier cleavage across the minor groove.

The antiparallel β-sheet that facilitates DNA binding forms a structure complimentary to the major groove of B-form DNA (Fig. 5). This motif facilitates the recognition of extended DNA sequences by maintaining a spacing of 6.9 Å between alternate side chains and by adopting a curvature and twist complementary to the DNA helical axis[16]. This structure is similar to, but shorter than, the paired β-sheets formed by the I-*Cre*I and PI-*Sce*I homing endonucleases[8,9]. The interactions formed in this interface represent both direct protein–DNA hydrogen bonds and water-mediated contacts, most of which are made by β4, using every alternate side chain in its sequence. This β-strand begins with Arg 61, which contacts the phosphate backbone near the scissile phosphate. As the β-strand continues to follow the major groove towards the end of the homing site, every alternate residue contacts the DNA. Gln 63 makes direct bipartite contact to Ade +6. Genetic studies demonstrate that adenine is essential at this position (R.M., unpublished observations). Lys 65 contacts Ade –8 and Gua –9 through water-mediated interactions. Finally, the Oγ of Thr 67 contacts the 5′ terminus of the oligonucleotide. Although the side chain of Arg 66 in this central β-strand points away from the DNA, the long side chain wraps around the β-strand to make two water-mediated contacts with P-9 and P-10. A small number of additional contacts are made in this region by the two additional β-strands that flank β4.

One base-specific contact in each half-site is made in the minor groove, where Lys 120 is bound to O2 of Thy 1 and makes a water-mediated contact to Ade 1. The A:T base pairs at these positions are strongly favoured, but can be reversed while still maintaining site

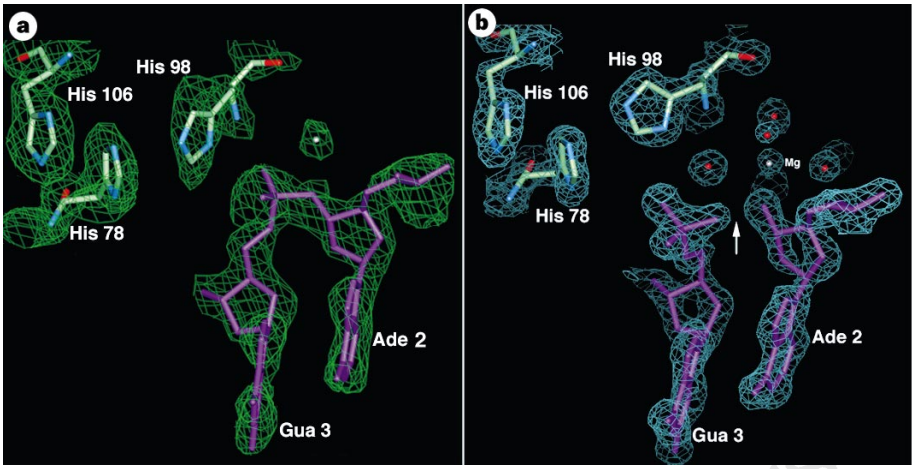

Figure 6 I-*Ppo*I homing endonuclease active site. **a**, $2F_o - F_c$ simulated annealing omit map (2.2 Å) of the substrate complex contoured at 1σ, showing the density around the phosphodiester backbone between Ade 2 and Gua 3. The distance from the Nδ nitrogen of His 98 to the scissile phosphate oxygen is 3 Å. **b**, A similar omit map (1.8 Å) of the cleaved product complex with a bound divalent cation. The 3′ hydroxyl oxygen leaving group is 4 Å from the phosphorus atom. The octahedral coordination of the bound cation is clearly visible in the map. The cation is directly coordinated by the 3′ oxygen of the hydroxyl leaving group and by the side-chain oxygen of Asn 119. The asparagine ligand to the metal is not shown for clarity, but density for its oxygen atom is visible behind and below the magnesium ion. Arg 61, which makes a 2.9 Å contact to the cleaved phosphate group, is also not shown. Anomalous difference maps calculated from crystals soaked with 10 mM manganese verify the location of the bound metal ion. This figure was made using QUANTA[24].

recognition by the endonuclease (R.M., unpublished observations). The second base pair at either side of the central cleavage site shows a similar dependence on A:T content for endonuclease activity. The apparent exclusion of G:C base pairs in the cleavage site may reflect the presence of a specific contact in the minor groove that would be difficult to form in the presence of the bulkier guanine nucleotide.

Endonucleases such as I-*Ppo*I that cleave across the minor groove of DNA (liberating 3′ overhangs) must cleave two phosphate groups separated by only 10–12 Å in an undistorted B-form double helix. The homing endonucleases I-*Cre*I and PI-*Sce*I have evolved an elegant structural solution to this problem: the conserved LAGLIDADG motif is used to generate a closely packed subunit or domain interface that appropriately positions active-site residues at the centre of the active site with minimal separation between them[8,9]. I-*Ppo*I uses a different strategy: it severely bends the homing-site DNA (Fig. 1), making the scissile phosphates more accessible for the two separate enzyme active sites.

At the point of the bend in each half-site, Leu 116 points into the minor groove and contacts Ade +2. Unlike the TATA-binding protein (TBP), which uses phenylalanine residues to intercalate between the bases of the TATA box[17,18], the I-*Ppo*I leucines are not large enough to insert between the base pairs. Instead, they make edge-on contacts with the Ade 2, unstacking these bases from their immediate neighbours (Gua 3). This strategy, and the resulting distortion of the DNA, is similar to that observed for the complex of the purine repressor PurR to its regulatory binding site[19]. Mutation of Leu 116 greatly reduces the catalytic efficiency of I-*Ppo*I (B. L. S., unpublished observations). The 55° bend across the central four base pairs of the homing site increases the average width of the minor groove in this region to 10 Å. The major groove across these same base pairs narrows to less than 5 Å, and the scissile phosphates are separated by 22 Å.

The strategy of DNA recognition and binding exhibited by the structure of the I-*Ppo*I homing-site complex agrees well with previous biochemical and genetic analyses of this endonuclease[5]. I-*Ppo*I requires at least 14 base pairs of its homing-site DNA sequence for efficient binding and cleavage, and it protects up to four base pairs to either side of the homing site in MPE–Fe(II) footprinting and protection analyses. A preference for purines flanking the homing site has been observed. This suggests that nonspecific contacts extending beyond the homing site may facilitate sequence-specific binding and cleavage. In the crystal structure, water-mediated base contacts and phosphate-backbone interactions extend beyond the direct protein–DNA contacts to generate an interface over 22 base pairs.

The structure of I-*Ppo*I complexed to its DNA substrate was determined under three separate conditions to characterize the active site (Fig. 6). The enzyme was initially crystallized with low concentrations of magnesium. Electron density maps of this structure indicate that the scissile phosphate group has been cleaved, generating a product complex. The divalent cation is bound in the active site, where it displays six-fold octahedral coordination to four water molecules, one protein side chain and one DNA oxygen. Crystals grown under these conditions were soaked for 1 h in 10 mM manganese, and difference Fourier maps were used to confirm the location of the bound metal cation. Finally, a cocrystal structure was determined in which the metal-bound oxygen atom of the DNA substrate was substituted with sulphur. This strategy inhibits cleavage in the crystal. The resulting structure is of an uncleaved substrate complex.

In the substrate complex, the conserved His 98 is within hydrogen-bond distance of the scissile phosphate group. This residue is an obvious candidate to act as a Lewis acid and stabilize the negative charge on the pentacoordinate phosphate transition state. There is no indication that a water bound to an adjacent phosphate group might act as a nucleophile. We prefer a model in which a metal-bound water (visualized in the cleaved product complex) is deprotonated by a general base and attacks the phosphate. In this structure, there is a weak peak, near the position of the bound metal ion in the product complex, that may be either a water molecule or a low-occupancy magnesium site.

In the product complex, the phosphate group is cleaved with the distance between the 3′ hydroxyl oxygen leaving group and the 5′ phosphorus atom refining to approximately 4 Å. The coordinations of the bound magnesium and manganese ions, determined separately, are identical. The metal ion is coordinated to the 3′ hydroxyl oxygen leaving group of the cleaved phosphate and to the side-chain oxygen of Asn 119 and has four visible coordinating water molecules, one of which directly contacts an oxygen atom of the cleaved 5′ phosphate group. The metal–oxygen bond distances range from 2.0 to 2.4 Å and form octahedral coordination spheres around the metals in both cases. Apart from the active-site divalent cation and

Table 1 Crystallographic data

Crystal complex	Nat_1 Zn, Mg (cleaved)	NAT_ALS Zn, Mg (cleaved)	Nat_Cd Cd, Mg (cleaved)	Nat_Mn Zn, Mn (cleaved)	Thio_DNA Zn, thio-DNA (uncleaved)	Hg (derivative 1)	IdU (2)
Internal merging and scaling							
Resolution (Å)	2.7	1.8	2.1	2.8	1.9	2.7	2.7
Reflections measured	50,053	362,336	83,203	85,277	179,752	60,392	33,285
Unique reflections	19,611	71,892	30,702	16,111	50,210	17,300	14,471
Redundancy	2.55	5.04	2.71	5.29	3.58	3.49	2.30
Completeness (%)	91.9	99.8	89.8	95.2	96.3	92.4	77.9
Average $I/\sigma(I)$	15.8	19.1	15.3	12.3	18.4	14.5	12.2
R_{sym}	0.045	0.049	0.048	0.074	0.033	0.053	0.054
Mosaicity (°)	0.30	0.37	0.37	0.60	0.22	0.40	0.25
MIR phasing							
Resolution (Å)						3.0	3.0
Mean isomorphous difference						4.7	5.7
R_{iso} (%)						9.2	11.7
R_{cullis}						0.87	0.73
Phasing power						0.78	1.45
Refinement							
Resolution (Å)		1.8	2.1	2.8	1.9		
R-factor		22.6	21.2	23.2	20.1		
R_{free}		24.7	30.4	30.7	24.2		
No. of protein atoms		2,490	2,490	2,490	2,490		
No. of nucleic acid atoms		856	856	856	854		
Total solvent molecules		403	547	8	363		
R.m.s. bond lengths (Å)		0.011	0.010	0.010	0.013		
R.m.s. angles (°)		1.7	1.4	1.8	1.5		
Ramachandran distribution (% core, allowed, generous, disallowed)		91.8, 7.8, 0.4, 0.0	90.2, 9.4, 0.4, 0.0	84.8, 13.7, 1.2, 0.4	92.2, 7.0, 0.8, 0.0		
Mean B-factors (protein/DNA, solvent)		15.5, 20.7	31.9, 53.0	24.0, 47.0	18.6, 23.12		

its coordinating Asn residue, the other groups within potential chemical-interaction distance of the cleaved phosphate group are His 98 and Arg 61. The largest structural differences in the active site produced by cleavage are the movement of the scissile phosphate towards Arg 61 and a rotation of the His 98 side chain, which maintains its contact to the phosphate group.

I-*Ppo*I displays a sequence that resembles a type II endonuclease signature (PD...(D/E)XK), formed by residues Pro 86–Asp 87... Asp 118–Asn 119–Lys 120. The residues identified in this motif are found in the crystal structure to not be involved in the enzyme active site. These residues are not conserved in a sequence alignment of the known His–Cys box endonucleases (Fig. 4). The endonuclease sequence signature identified in type II restriction endonucleases is only coincidentally present in I-*Ppo*I (and probably in many other protein sequences), and does not serve an important role in endonuclease structure or function.

In summary, the long homing-site interface formed by the small I-*Ppo*I endonuclease is facilitated both by an extended fold dependent on bound metals that largely substitute for a compact hydrophobic core, and by sparsely distributed protein–DNA contacts. Economy of structure appears to be an important feature of these proteins, which are probably limited in their overall size by the location of their open reading frames inside functional introns. Comparison of the structures of restriction and homing endonucleases provides a detailed view of how these proteins have independently evolved solutions to the problem of site-specific DNA recognition and cleavage. □

Methods

Crystals were of I-*Ppo*I complexed with a 20-base-pair synthetic DNA duplex (Oligos, Etc.) containing a palindromic variant of the wild-type I-*Ppo*I homing site[20]. Crystals grown using this protocol contain a stable, cleaved product complex, with a single bound magnesium ion in each active site. The bound metal position was confirmed by soaking 10 mM manganese into crystals and calculating an anomolous difference Fourier map. Crystals of an uncleaved substrate complex were prepared in an identical manner, except that the hydroxyl oxygen of the 3′ leaving group in the DNA oligonucleotide was substituted with a sulphur atom. All data, except set NAT_ALS (Table 1), were collected at 100 K on an RAXIS II-C area detector mounted on a Rigaku RU200 rotating anode X-ray generator (Molecular Structure Corporation). Data set NAT_ALS was collected on an ADSC 4 panel CCD area detector at beamline 5.0.2 at the Advanced Light Source (LBNL, Berkeley) using a wavelength of 1.0 Å. All RAXIS data were reduced using the DENZO/SCALEPACK crystallographic data reduction package[21]; the CCD data were reduced using MOSFLM 5.5 (ref. 22). One isomorphous derivative was prepared by synthetic incorporation of an iododeoxyuridine in the DNA duplex. The second was prepared by a 1-h soak of the native crystals in a 1 mM $Hg(CN)_2$ solution. All initial multiple isomorphous replacement phase calculations and phase refinements were done using the 'Native 1' data set in the CCP4 package[23] with phases calculated using the program MLPHARE. Phases calculated at 3.50 Å resolution using only these data yielded an interpretable electron density map that was further improved by solvent flattening, histogram matching and phase extension to 2.8 Å resolution. These phases were applied to the higher quality 'Cd' data set for model building, and extended to 2.1 Å resolution after the initial chain trace. Phase-combined maps were calculated using SFALL and SIGMAA to resolve ambiguous regions of the original DM electron density map. There is a single I-*Ppo*I dimer and DNA duplex in the asymmetric unit of the crystal. Model building was carried out using QUANTA[24], followed by refinement using XPLOR[25]. The higher-resolution NAT_ALS data set then became available for final refinement of the cleaved product complex to 1.8 Å resolution. The stereochemical quality of the protein model was examined throughout the refinement using PROCHECK[26]. The final models contain no residues with unfavourable backbone dihedral angles with over 90% of the residues in the core region of the Ramachandran plot.

Received 19 January; accepted 8 April 1998.

1. Belfort, M. & Roberts, R. Homing endonucleases—keeping the house in order. *Nucleic Acids Res.* **25,** 3379–3388 (1997).
2. Belfort, M., Reaban, M. E., Coetzee, T. & Dalgaard, J. Z. Prokaryotic introns and inteins: a panoply of form and function. *J. Bacteriol.* **177,** 3897–3903 (1995).
3. Belfort, M. & Perlman, P. S. Mechanisms of intron mobility. *J. Biol. Chem.* **270,** 30237–30240 (1995).
4. Cooper, A. A. & Stevens, T. H. Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci.* **20,** 351–356 (1995).
5. Ellison, E. L. & Vogt, V. M. Interaction of the intron-encoded mobility endonuclease I-*Ppo*I with its target site. *Mol. Cell. Biol.* **13,** 7531–7539 (1993).
6. Muscarella, D. E. & Vogt, V. M. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **56,** 443–454 (1989).
7. Johansen, S., Embley, T. M. & Willassen, N. P. A family of nuclear homing endonucleases. *Nucleic Acids Res.* **21,** 4405 (1993).
8. Duan, X., Gimble, F. & Quiocho, F. Crystal structure of PI-*Sce*I, a homing endonuclease with protein splicing activity. *Cell* **89,** 555–564 (1997).
9. Heath, P. J., Stephens, K. M., Monnat, R. J. & Stoddard, B. L. The structure of I-*Cre*I, a group I intron-encoded homing endonuclease. *Nature Struct. Biol.* **4,** 468–476 (1997).

10. South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30,** 6342–6349 (1991).
11. Luisi, B. F. *et al*. Crystallographic analysis of the interaction of the glucorticoid receptor with DNA. *Nature* **352,** 497–505 (1991).
12. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356,** 408–414 (1992).
13. Everett, R. D. *et al*. A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J. Mol. Biol.* **234,** 1038–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C_3HC_4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237,** 201–211 (1994).
15. Borden, K. L. B. *et al*. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14,** 1532–1541 (1995).
16. Phillips, S. E. V. The β-ribbon DNA recognition motif. *Ann. Rev. Biophys. Biomol. Struct.* **23,** 671–701 (1994).
17. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365,** 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. *Nature* **365,** 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266,** 763–770 (1994).
20. Flick, K. E. *et al*. Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6,** 1–4 (1997).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
22. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).
23. CCP4 *The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1979).
24. QUANTA96 *X-ray Structure Analysis User's Reference* (Molecular Simulations, San Diego, 1996).
25. Brünger, A. *XPLOR version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 383–290 (1993).
27. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11,** 134–138 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBL laboratories), beamline 5.0.2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. K.E.F. was supported by an NIH training grant and the American Heart Associaiton. M.S.J. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.fhcrc.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos lipp, 1a73, 1a74).

corrections

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature **390,** 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeannine D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D'Andrea, Cheryl Bowman, Claire Fujii, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature **390,** 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (*cysC*) is not needed in the pathway as outlined, as adenylyl sulphate reductase (*aprAB*) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (*sat*), adenylylsulphate reductase (*aprAB*), and sulphite reductase (*dsrABD*). We thank Jens-Dirk Schwenn for bringing this error to our attention. □