The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif

Lei Zhao^{1,2}, Richard P Bonocora^{3,4},
David A Shub³ and Barry L Stoddard^{2,*}

¹Graduate Program in Molecular Biophysics, Structure and Design, University of Washington, Seattle, WA, USA, ²Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, USA and ³Department of Biological Sciences and Center for Molecular Genetics, University at Albany, State University of New York, Albany, NY, USA

The homing endonuclease I-Ssp6803I causes the insertion of a group I intron into a bacterial tRNA gene—the only example of an invasive mobile intron within a bacterial genome. Using a computational fold prediction, mutagenic screen and crystal structure determination, we demonstrate that this protein is a tetrameric PD-(D/E)-XK endonuclease—a fold normally used to protect a bacterial genome from invading DNA through the action of restriction endonucleases. I-Ssp6803I uses its tetrameric assembly to promote recognition of a single long target site, whereas restriction endonuclease tetramers facilitate cooperative binding and cleavage of two short sites. The limited use of the PD-(D/E)-XK nucleases by mobile introns stands in contrast to their frequent use of LAGLIDADG and HNH endonucleases—which in turn, are rarely incorporated into restriction/modification systems.

The EMBO Journal (2007) **26**, 2432–2442. doi:10.1038/sj.emboj.7601672; Published online 5 April 2007

Subject Categories: microbiology & pathogens; structural biology

Keywords: group I intron; homing endonuclease; PD-(D/E)-XK fold; protein–DNA binding; restriction endonuclease

Introduction

Homing endonucleases are highly specific DNA binding and cleavage enzymes that are encoded as open reading frames (ORFs) embedded within introns or inteins (Lambowitz and Belfort, 1993; Mueller *et al*, 1993; Belfort and Perlman, 1995a; Belfort *et al*, 1995b; Belfort and Roberts, 1997; Jurica and Stoddard, 1999; Chevalier and Stoddard, 2001b). Homing endonucleases promote the mobility of the intervening sequences (and their own reading frames) by generating

double-strand breaks in homologous alleles that lack the intron or intein. Break repair leads to transfer of the element via homologous recombination, using the allele that contains the homing endonuclease gene (HEG) as a template. In rare cases, homing endonucleases can also be encoded by free-standing genes, with their mobility accomplished by a similar mechanism that operates independently of the presence of a surrounding intervening sequence (Belle *et al*, 2002). In either case, HEGs are selfish DNA sequences that are inherited in a dominant, non-Mendelian manner. In addition to vertical transmission, homing endonucleases also drive the ectopic horizontal transfer of introns and inteins, and encourage their persistence in prokaryotic and phage genomes.

Homing endonucleases from group I introns tend to be small proteins of less than 30 kDa per peptide chain, owing to size constraints imposed by their surrounding host sequences, which must fold into structures capable of splicing. Despite their small size, these proteins recognize long DNA target (typically 14–40 bp), which reduces their toxicity in the host organism. Within these long targets, homing endonucleases tolerate sequence variation at individual base pairs; this attenuated fidelity enables them to adapt to sequence drift in their target and increases their potential for ectopic transfer.

There are five known homing endonuclease families, each of which has arrived at the optimal balance of protein size, DNA-binding specificity and attenuated fidelity that is most suitable for evolutionary success in their host genomes. Homing endonuclease families are classified and named according to their most conserved sequence and structural motifs, and generally are localized to distinct biological and genomic niches (Stoddard, 2005).

The largest homing endonuclease family, termed ‘LAGLIDADG’, is usually encoded in mitochondrial or chloroplast genomes in single-cell eukaryotes or in archaea (Heath *et al*, 1997; Jurica *et al*, 1998; Chevalier *et al*, 2001a, 2004; Bolduc *et al*, 2003; Spiegel *et al*, 2006). In contrast, the ‘His-Cys box’ and ‘HNH’ endonucleases are found in protists and phage, respectively. They share similar active sites and appear to be descended from a common ancestor; however, their surrounding tertiary structures have diverged greatly (Friedhoff *et al*, 1996; Flick *et al*, 1998; Kuhlmann *et al*, 1999; Shen *et al*, 2004). A fourth family, named ‘GIY-YIG’ endonucleases, is also encoded within phages as well as in organellar genomes (Kowalski *et al*, 1999).

A fifth type of homing endonuclease is encoded within group I introns in bacterial genomes (Xu *et al*, 1990; Reinhold-Hurek and Shub, 1992; Bonocora and Shub, 2001). The majority of these bacterial introns, which are few in number, are found in tRNA genes of formylmethionine, leucine, isoleucine and arginine (Edgell *et al*, 2000; Nesbo and Doolittle, 2003). Although most of these introns are devoid of HEGs, a protein-coding ORF was found in the

*Corresponding author. Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N. A3-025, Seattle, WA 98109, USA. Tel.: +1 206 667 4031; Fax: +1 206 667 3331; E-mail: bstoddard@fhcrc.org

⁴Present address: Gene Expression and Regulation Section, Laboratory of Molecular and Cellular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892-0830 USA

Received: 8 February 2007; accepted: 9 March 2007; published online: 5 April 2007

tRNA^{fMet} gene of *Synechocystis* strain 6803 (Bonocora and Shub, 2001). This protein, named 'I-Ssp6803I' (hereafter called 'I-SspI' for simplicity), and its presumed relatives are believed to be involved in the persistence of closely related group I introns that are found at the same genomic insertion point in several cyanobacterial genera (Biniszkiwicz *et al*, 1994; Paquin *et al*, 1997).

The ORF encoding I-SspI contains 150 codons and is the smallest HEG characterized to date. The I-SspI ORF also forms part of the P1 stem, and the entire P2 stem-loop of its host intron (Bonocora and Shub, 2001). The involvement of the endonuclease start codon in the secondary structure of the intron, and the absence of a recognizable ribosomal binding site, indicate that the reading frame may be inefficiently translated in its natural host, perhaps as a mechanism to reduce toxicity.

I-SspI harbors little sequence identity with other known nucleases or functionally annotated reading frames. Previously described biochemical studies (Bonocora and Shub, 2001), combined with the structural analysis reported here, indicate that the DNA target site is approximately 23 bp in length, corresponding to a pseudopalindrome (5'-TCGTCGGGCTCATAACCCGAAGG-3'). This site spans the sequence encoding the anticodon loop in the intron-minus tRNA^{fMet} gene (the bases in boldface indicate positions of palindromic symmetry; bases underlined indicate the fMet anticodon). I-SspI cleavage produces complementary 3-base, 3' overhangs (5'-CAT-3' and 5'-ATG-3') that exactly flank the fMet anticodon (Biniszkiwicz *et al*, 1994; Bonocora and Shub, 2001). Biochemical studies of I-SspI have been impeded by toxicity of the wild-type enzyme (probably due to cleavage of *Escherichia coli* tRNA^{fMet} genes, which are highly conserved across the enzyme's target site).

To characterize the protein factor that appears to be responsible for intron persistence in bacteria, we have combined the use of a consensus computational fold prediction, a mutagenic screen for inactivating mutations in its active site, and X-ray crystallography. The results provide the first example of the use of the PD-(D/E)-XK protein fold, which is most commonly associated with defense of the bacterial genome by restriction endonuclease, for the purpose of intron invasion and persistence. This study provides insight into the balance of structural constraints that dictate the relative success and penetrance of this nuclease fold into the different processes of restriction and homing.

Results and discussion

Identification of the I-SspI active site and the PD-(D/E)-XK nuclease fold

Attempts to clone the toxic wild-type I-SspI ORF into a bacterial expression system were unsuccessful. In order to produce protein for crystallization trials (as well as locate active site residues), we decided to identify a catalytically inactive mutant which could be overexpressed in *E. coli* without compromising its stability or DNA-binding affinity.

A wild-type I-SspI reading frame was synthesized (Supplementary data) and subcloned in a promoterless pUC vector for plasmid amplification (Blue Heron Inc.). This gene was then subcloned into a pET15b expression vector (Novagen Inc.) under a combination of bacterial strains and media conditions that repress transcription. Transformation

of this vector into expression strains, including high stringency hosts such as *BL21(DE3)pLysS* (Novagen), did not produce colonies. Therefore, we decided to use the results of such a transformation as the basis for a screen for inactivating mutations in the endonuclease active site.

We designed a mutagenic strategy to focus on the most ubiquitous known catalytic property of endonucleases: binding of divalent cations in the active site that are required during the reaction (Yang *et al*, 2006). Reasoning that metal ions are most often bound by aspartate (and less frequently by glutamate) residues, we designed an 'Asp to Ala' scan protocol. Primers were designed to mutate each aspartate in the I-SspI ORF (nine positions total). These primers were combined into a single 'multichange' mutagenesis PCR (Stratagene Inc.) and the resulting mixture of mutagenic products was directly transformed into the BL21(DE3) *E. coli* expression strain. The plasmid-encoded I-SspI ORFs from individual colonies were sequenced, and we determined that mutation of a single aspartate residue (D8A) permitted bacterial growth (Supplementary Table S1). We therefore reasoned that Asp 8 is very likely a catalytic residue, although a structural role could not be ruled out.

A subsequent sequence-based search by PSI-BLAST, as well as analyses with sequence/structure threading servers such as 3D-PSSM or Phyre (Kelley *et al*, 2000), failed to reveal any homologues with known function, although a free-standing hypothetical reading frame was detected in T7 phage (gene 5.3). However, a structure-based sequence comparison server (meta server, <http://bioinfo.pl/meta/>) indicated a weak match of the first 100 residues of I-SspI against PDB entry 1GEF, with overall 18% sequence identity and a Z-score of 31.5. This structure corresponds to a Holliday junction resolvase (Hjc) from the archaea *Pyrococcus furiosus*, that contains the 'PD-(D/E)-XK' core fold found in most type II restriction endonucleases (Nishino *et al*, 2001). A sequence alignment with archaeal resolvase enzymes (Figure 1) allowed us to create a homology model of the I-SspI N-terminal core fold. At the time that this study was initially submitted for publication, this structural fold prediction was also described by another group, using the same computational server (Orlowski *et al*, 2007); a comparison of the structure prediction with the crystal structure is described below.

A sequence comparison of I-SspI with the Hjc resolvase family (Komori *et al*, 2000) revealed conservation of several active site residues: Glu 9 in Hjc-Pfu (Glu 11 in I-SspI), Asp 33 (Asp 36), Glu 47 (Gln 49) and Lys 49 (Lys 51). Although Asp 8 is not conserved in this alignment, it is only three residues away from the conserved Glu 11, on the same side of a α -helix in the active site. A mutant construct of I-SspI corresponding to Glu 11 to Gln (E11Q) was also successfully transformed and overexpressed in *E. coli*. This led us to hypothesize that I-SspI may contain a canonical PD-(D/E)-XK fold with Asp 8 and Glu 11 both involved in the structure and function of the endonuclease active site.

Structure determination

Although catalytically inactive point mutants (D8A or E11Q) are overexpressed in *E. coli*, the majority of the protein is insoluble. On the basis of our model of the enzyme core described above, a phenylalanine residue (F55) was predicted to be solvent-exposed and far removed from the DNA-binding surface (Supplementary Figure S1). Incorporation of a lysine

Figure 1 Sequence alignment between the bacterial I-Ssp6803I homing endonuclease and archaeal Holliday junction resolvases. Only the first 110 residues of I-SspI, that align well, are shown with the homologous regions of the Hjc sequences. The final 40 residues of I-Ssp6803I that are not shown to participate in structural elaborations on the PD-(D/E)-XK core fold that are unique to I-SspI. Secondary structure elements of the homing endonuclease are shown above the alignment; structural elements from *Pyrococcus furiosus* Hjc are shown below. All of these elements are conserved in I-SspI with the exception of α -helix 2 ($\alpha 2$), which instead is an extended loop ('L1' in the text and subsequent figures) that contacts the DNA target site. The blue stars above the alignment indicate conserved residues at active sites of the Hjc resolvase family. Residue labels in parentheses above the sequence alignment indicate mutations present in the crystal structure, as described in the text. Sequence alignments were carried out by ESPript (Gouet *et al*, 1999).

Table 1 Crystallographic data

Crystal complex	Native (E11Q/F55K)	Se-Met_peak (E11Q/F55K/L16M/L21M)	Se-Met_remote (E11Q/F55K/L16M/L21M)
<i>Crystallographic data</i>			
Resolution (Å)	4.0	3.1	3.3
Wave length (Å)	1.5428	0.979 (peak)	0.965 (remote)
Space group	I422	I422	I422
Cell parameters (Å)	$a = b = 143.09$, $c = 319.43$, $\alpha = \beta = \gamma = 90^\circ$	$a = b = 143.78$, $c = 319.18$, $\alpha = \beta = \gamma = 90^\circ$	$a = b = 143.96$, $c = 319.83$, $\alpha = \beta = \gamma = 90^\circ$
Completeness (%)	99.9 (100.0)	94.0 (61.4)	100.0 (100.0)
R_{sym} (%)	13.1 (45.1)	8.3 (37.4)	15.2 (63.0)
I/σ	21.8 (6.3)	20.0 (3.5)	16.2 (3.5)
Redundancy	14.2 (14.6)	7.6 (7.6)	9.5 (9.4)
<i>MAD phasing</i>			
Resolution (Å)		3.3	
Phasing power		1.035	
<i>Refinement</i>			
Resolution (Å)		3.1	
R-factor (%)		26.6	
R-free (%)		31.3	
R.m.s. bond length (Å)		0.008	
R.m.s. bond angle (deg)		1.995	
Ramachandran distribution		83.1% core 13.3% allowed 1.8% generous 1.8% disallowed	
Mean B-factor (Å ²)		75.10	

residue at this position (to create a E11Q/F55K double mutant) allowed the preparation of milligram quantities of highly pure, easily concentrated material. In addition, a pair of leucine residues (L16 and L21) were changed to methionine to facilitate phasing.

The resulting protein construct displays nanomolar affinity to its DNA target site using isothermal titration calorimetry (Supplementary Figure S2) and was cocrystallized bound to a 27 bp DNA duplex containing an I-SspI target site. The structure of the complex was determined using the

multiwavelength anomalous dispersion (MAD) method, with data collected at beamline 5.0.2 at the Advanced Light Source (ALS) synchrotron. The experimental electron density map was of excellent quality (Supplementary Figure S3). The structure was refined to 3.1 Å resolution with $R_{\text{work}}/R_{\text{free}} = 0.266/0.313$ (Table I).

Overall quaternary protein structure and stoichiometry of DNA binding

The structure of the I-SspI/DNA complex consists of one protein tetramer bound to a single DNA duplex; the crystallographic asymmetric unit contains one copy of this complex (Figure 2). We were able to model completely the entire chain of both DNA-bound monomers and the entire DNA molecule. The unbound monomers were also easily modeled, except for a short disordered surface loop region in each subunit (residues 71–82 in monomer C and residues 68–82 in monomer D) that is only ordered upon DNA binding.

Several independent lines of evidence agree with the observed protein:DNA stoichiometry: (i) the protein runs as a tetramer on a size exclusion column; (ii) the crystals were grown in a three-fold excess of DNA relative to the protein tetramer (therefore, potential binding at the second site was not limited by the DNA concentration) and (iii) binding experiments using isothermal titration calorimetry clearly indicates a DNA:protein stoichiometry that agrees with the crystal structure (Supplementary Figure S2). As discussed below, binding of a single DNA duplex induces a rearrangement in the packing of the tetramer that prevents binding of a second site.

The I-SspI PD-(D/E)-XK fold

Each I-SspI monomer displays a topology containing four α -helices and nine β -strands (Figure 3A). The core catalytic region consists of one α -helix ($\alpha 1$) surrounded by five β -strands ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 7$ and $\beta 8$). Three of these elements ($\alpha 1$, $\beta 1$ and $\beta 2$) are involved in assembly of the protein tetramer. Two additional α helices ($\alpha 3$ and $\alpha 4$) pack against this core fold and comprise the C-terminal end of the monomer.

Visual examination and analyses using the DALI structure comparison server (Holm and Sander, 1996) indicate that this domain corresponds to the canonical PD-(D/E)-XK nuclease fold, found in most restriction endonucleases. In this fold, the β -sheet is concave and markedly curved toward the $\alpha 1$ -helix. All residues thought to be important for catalysis by I-SspI and its immediate structural homologues are positioned at the concave side of the five β -strands, at one end of each subunit.

In addition to restriction endonucleases, the PD-(D/E)-XK fold has also been observed in other enzymes involved in DNA rearrangements and modifications, including phage exonucleases, archaeal Holliday junction resolvases, phage T7 endonuclease I, transposase TnsA and certain DNA repair enzymes such as MutH and Vsr (Bujnicki *et al*, 2001). Structural comparisons with previously determined crystal structures using the DALI server reveals that the overall structure of the I-SspI monomer is most similar to the archaeal Holliday junction resolvases, typified by the Hjc enzyme from *Pyrococcus furiosus*, with a Z-score of 9.9 and r.m.s.d. for aligned C α atoms of 2.4 Å (1.9 Å across the catalytic core) (Figure 3B). Whereas resolvase enzymes recognize a specific DNA backbone conformation without any strong sequence

Figure 2 Structure of I-SspI bound to its DNA target site. Protein subunits are each colored separately. The complex is shown in three mutually orthogonal orientations in (A–C). The buried surface area in each subunit interface is indicated. Two bound calcium ions are shown as red spheres. Loop L1 from monomers A and B is indicated by double black arrows. These loops are primarily associated with the central bases of the target site. These two loops do not interact with the bases in an identical manner—reflecting the asymmetry of this region of an otherwise symmetric target site. The same loops are disordered in subunits C and D, which are not bound to DNA and display minimal subunit contacts. The crystallization oligonucleotide construct is shown below panel A. The cleavage sites are indicated by cyan triangles. The base positions corresponding to physiological homing site are shown in red and the central 3-bases (corresponding both to the 3' overhangs produced by cleavage and to anticodon triplet for fMet) are bold. The palindromic base pairs in the structure are underlined.

preference (Komori *et al*, 2000; Nishino *et al*, 2001), I-SspI recognizes a long DNA target sequence. This difference in binding activities results from unique structural elaborations on the core endonuclease fold as discussed below.

Of the type II restriction endonucleases that have been visualized to date, the closest structural homologue of I-SspI is *PvuII* (Figure 3C), with a DALI Z-score of 5.6 and an r.m.s.d. over the aligned C α atoms of 3.3 Å. However, the I-SspI endonuclease (which recognizes a 23 bp target) is smaller (150 residues) than *PvuII* (157 residues), which recognizes a 6 bp target sequence. This suggests that the

Figure 3 Structural comparison of protein subunits from the I-Ssp6803I homing endonuclease, the Hjc Holliday junction resolvase and the *PvuII* restriction endonuclease. (A) Structure and topology diagram of a single homing endonuclease subunit. The secondary structural elements are labeled and colored as follows: the PD-(D/E)-XK catalytic core region is pink and peripheral elaborations on that core are green. The N- and C-terminal residues of the secondary structural elements are indicated in the topology diagram. Catalytic residues are shown as sticks in the model on the left and labeled in red on the right. Regions involved in DNA recognition are indicated by dotted boxes and are numbered as shown in Figure 6 and described in the text. (B) The Hjc Holliday junction resolvase subunit. This structure has not been determined in the presence of DNA. (C) The *PvuII* restriction endonuclease subunit. Inlay: superposition of the I-SspI and Hjc catalytic cores (r.m.s.d. 1.9 Å).

structural elaborations to a PD-(D/E)-XK domain required for recognition of a long DNA target with reduced fidelity can be achieved at least as economically as the alternative elaborations required for recognition of a short site with absolute fidelity. As approximately the same number of direct contacts to DNA base pairs are made by I-SspI and *PvuII*, we also suggest that restriction endonucleases require additional protein mass primarily to expand their surface complementarity to the phosphoribosyl backbone and/or induce significant DNA bending, as strategies to increase fidelity.

Assembly of the endonuclease tetramer

The protein tetramer measures approximately $80 \times 80 \times 40$ Å and displays 222 (D2) symmetry that is broken by DNA binding across two of the four protein subunits (Figure 2). The catalytic cores of the four subunits show nearly identical structures, with an average r.m.s.d. value between subunits of 1 Å. Two of the protein monomers (A and B) interact with the DNA, which is uncleaved and slightly bent by $\sim 25^\circ$ around its central base. Approximately 3500 Å^2 are buried in the binding interface between protein and DNA.

Two additional protein monomers (C and D) complete the tetramer and point in the opposite direction from the protein-DNA complex, in a back-to-back arrangement with a nearly 90° rotation (Figure 2). Although the core folds of the individual subunits are closely superimposable to each other, the relative orientation of the two DNA-bound subunits differ from that of the unbound subunits. Superposition of the

DNA-bound subunits against their unbound counterparts (Figure 4) indicates that this difference consists of a rigid-body rotation of protein subunits by approximately 5° .

The tetrameric architecture of I-SspI is maintained by three pairs of unique packing arrangements between DNA-bound and DNA-unbound monomers, generating a dimer-of-dimers in which two dimers each bind to a DNA half-site, and the two DNA-bound subunits (and their symmetry mates) are in looser contact with one another (Figure 2). A total of approximately 3300 Å^2 of protein surface area is buried within the tetramer.

Monomers A-D (and B and C), are associated with one another through an antiparallel packing of their $\alpha 1$ helices from the PD-(D/E)-XK fold, creating two interfaces that each bury approximately 500 Å^2 (Figure 2B). This interaction is mediated by van der Waals interactions between small hydrophobic residues presented by one side of each helix. In contrast, this same helix is exposed to solvent in the homodimeric Hjc resolvase and is populated by highly polar and charged residues. This comparison demonstrates the structural differences that evolve as members of a protein fold family diverge from a common ancestor, resulting in different quaternary structures.

In addition, monomers A-C (and B-D) form a two four-stranded β -sheets at their interfaces (Figure 2C), again using secondary structure from the nuclease core fold. Two β -strands from each monomer ($\beta 1$ and $\beta 2$) participate in these interfaces, that each bury approximately 650 Å^2 .

Figure 4 Superposition of DNA-bound and unbound subunits in I-SspI. The endonuclease subunits are colored as in Figure 2. The DNA-bound subunits (A and B) and their bound DNA ligand are superimposed on subunit (C) of the two DNA-free monomers. As discussed in the text, this analysis indicates that the unbound subunits display a rigid-body rotational difference in their relative orientations and packing, as compared to the DNA-bound subunits. This results in a difference of approximately 6 Å in the position of the DNA-binding surface of subunit D relative to subunit A. The observed structure and packing of the C/D subunits cannot be accommodated for DNA binding either by DNA straightening (because of steric crowding in the central minor groove) or by repacking of subunit D (which would destabilize the A/D dimer interface).

Between these two sets of dimer interactions, none of which involve the interface between the DNA-bound subunits, approximately 2300 Å² of protein surface area is buried.

The packing described above generates contacts between surface loops (L1 and L1') from each of the DNA-bound protein subunits (Figure 2A). Approximately 400 Å² of surface area is buried between these two loops, which contact the major groove of the DNA target site underneath the central 3 bp (positions −1, 0 and +1) of the target site. In the DNA interface, one L1 loop is closely associated with the DNA target, whereas the other is more distant. This asymmetry is caused by the corresponding asymmetry of the target site's three central base pairs, which provide a superior binding target for the L1 loop in one direction (5'-CAT-3') than in the opposing direction (5'-ATG-3'). In the opposing protein subunits (C and D), the same loops are largely disordered, leaving only a pair of side-chain contacts between the cores of those subunits.

Binding of DNA targets by endonuclease tetramers

Superposition of the DNA-bound subunits and their DNA target against the unbound subunits (Figure 4), indicates that neither the DNA nor the protein tetramer can be remodeled to allow binding of a second target site without either (i) imposing unreasonable steric crowding in the central minor groove, or (ii) destabilizing the protein tetramer. It therefore appears that binding of the first DNA duplex to protein subunits A and B breaks perfect 222 symmetry in the unbound enzyme, and induces a movement subunits relative to one another that is incompatible with high-affinity binding of a second duplex.

A tetrameric enzyme assembly is also generated by many type II restriction endonucleases, and has been described in crystallographic structure analyses of DNA-bound complexes of SfiI (Vanamee *et al*, 2005) and NgoMIV (Deibert *et al*, 2000). Such quaternary structures appear to have often evolved in restriction endonucleases for the purpose of establishing mechanism that requires the presence and bind-

ing of two cognate recognition sites for efficient cleavage, through positive cooperativity and allosteric activation (Gowers *et al*, 2004). Such behaviors can lead either to enhanced cleavage of one of the bound target sites (a type IIe restriction mechanism) or of both bound sites in a coordinated manner (a type IIlf mechanism, displayed by SfiI and NgoMIV). This behavior may be important to avoid undesirable cleavage of spontaneously demethylated bacterial host sites.

The use of a tetrameric assembly by the homing endonuclease I-SspI appears to facilitate binding of a single long target site by allowing the core PD-(D/E)-XK domains to be far apart. In contrast, tetramer assembly by SfiI facilitates binding of two short DNA sites as described above, with the catalytic cores of each functional dimer packed more closely together. These different purposes for tetramer formation are reflected in very different arrangements of the endonuclease subunits for I-SspI and SfiI, each of which conforms to D2 (222) symmetry (Figure 5). Head-to-head stacking of the two DNA-bound dimers in the restriction endonuclease places the PD-(D/E)-XK domains, their DNA-binding surfaces and the active sites close together in the individual protein-DNA interfaces—a necessary property for recognition of a short target site. In contrast, 'side-by-side' packing of the enzyme subunits in I-SspI allows the DNA-bound nuclease cores to be much more loosely associated with one another and to then distribute additional DNA-binding surfaces (displayed as elaborations that project from the core PD-(D/E)-XK fold) toward the distal ends of the longer target site. The use of a tetrameric endonuclease assembly specifically to stabilize the functional dimer on the DNA target was described originally for the SfiI-DNA structure (Vanamee *et al*, 2005).

DNA target recognition

The physiological DNA homing site is a pseudopalindrome (5'-TCGTCGGGCTCATAACCCGAAGG-3'), with the sequence differing between DNA half-sites at four base pairs: ±11, ±9, ±3, and ±1. In addition, a single A:T base pair (position '0')

Figure 5 Topology of I-Ssp6803I tetramer assembly and comparison with the SfiI restriction endonuclease. (A) Active sites and overall tetrameric packing of the I-Ssp6803I homing endonuclease. (B) Active sites and overall tetrameric packing of the SfiI restriction endonuclease. These two endonuclease have the same core fold and similar cleavage patterns (producing complementary 3-base, 3' overhangs, as a result of cleavage across the minor groove) as shown below the models. Active sites from different monomers are colored in cyan or green (only secondary structures carrying the active site residues are shown for clarity). The cleavage sites on the DNA are indicated by red spots. The general architecture of the tetrameric assembly is indicated by the cartoon blocks representation, and are colored according to the corresponding protein subunits. The ribbon diagrams are shown with the 'B' subunit from each structure in roughly similar orientations, to facilitate direct comparison of the tetrameric packing of the endonucleases.

is located at the exact center of the target, and also breaks symmetry in the site and its protein-bound complex. Biochemical DNA protection assays indicated that bases +9 and +11 are bound more tightly than are their counterparts at -9 and -11 (leading to significant differences between these bases in footprinting experiments). Therefore, in the DNA construct used for crystallization the base pairs at positions -9 and -11 were converted to match their symmetry mates (Figure 2).

Overall, the DNA displays a slightly bent B-form conformation that curves away from the protein. The minor groove at the center of the DNA is significantly broadened to ~15 Å. Four discontinuous elaborations that extend from the PD-(DE)-XK protein fold are largely responsible for DNA recognition and binding, and make a variety of contacts across the entire length of the target. The protein regions involved in DNA contacts are numbered and labeled (1 through 4) in Figures 3 and 6 and correspond to the description below.

First, the N-terminal ends of helices $\alpha 1$ and $\alpha 1'$ contact the central region of the DNA (base pairs -2 to +2), where they

insert into the minor groove and contribute two residues to each of the active sites. Second, the L1 surface loop from subunit 'A' wraps around the DNA and contacts the opposite edge of the same base pairs, making contacts in the major groove at base pairs -1, 0 and +2. This interaction is asymmetric, as the same loop from subunit 'B' is not in contact with the DNA: this difference is a result of the corresponding asymmetry across the center of the DNA target.

Third, two short antiparallel β -sheets ($\beta 5$ - $\beta 6$ and $\beta 4$ - $\beta 9$) are arranged end to end (in tandem) and provide additional contacts in the major groove of each half-site, from base pair positions 3-7. Fourth and finally, another protein surface loop (L2) extends from the end of strand $\beta 9$ and makes contacts to the most distal ends of the DNA half-sites, within the minor groove at base pairs ± 10 and 11.

Thus, a discontinuous pair of short antiparallel β -sheets (consisting of two strands each, arranged end to end), along with two surface loops, wrap around 23 contiguous base pairs of the DNA target site and establish a mixture of contacts to the bases and to backbone atoms. The strategy of using β -strands for DNA target recognition is somewhat reminiscent of that used by other families of homing endonucleases such as the LAGLIDADG enzymes. However, in LAGLIDADG endonucleases the β -sheet DNA-binding platform is a single continuous structure in each protein domain that is an intimate part of the overall protein fold. In contrast, the constraints imposed by the use of discontinuous, surface-exposed elaborations on the PD-(D/E)-XK nuclease fold of I-SspI appears to reduce the density of contact side chains in the interface.

As is observed in other homing endonuclease-DNA co-crystal structures, the number of contacts to individual base pairs is variable and undersaturated (Figure 6). At least 12 direct hydrogen bond contacts are made between protein side chains and DNA bases in the major groove of each half-site, corresponding to contacts to approximately one-third of the available hydrogen bond contact points in the DNA major groove. In addition, the protein makes extensive contacts in the minor groove to the central A:T base pair, and a pair of additional minor groove contacts to individual N2 nitrogens of the cytosines at positions ± 10 and 11 (at the distal ends of the target site).

The target site displays the highest density of protein contacts across its central 13 bp (-6 to +6), whereas the flanking positions (± 7 through 11) exhibit fewer direct contacts. In the structure of the tRNA^{fMet}, the bases encoded by the central 13 positions in the DNA target site correspond to the majority of the anticodon stem loop, whereas several bases encoded by the flanking sequence are unpaired in the tRNA. In particular, bases ± 4 through 6 of the I-SspI target site encode a run of three consecutive G:C base pairs in the tRNA anticodon stem, which is a sequence feature that is diagnostic of tRNA^{fMet}. Thus, the contacts made by the homing endonuclease (and its likely pattern of specificity) appears to mirror the sequence constraints of the host gene and its tRNA product.

Active-site architecture

Phosphodiester bond hydrolysis by a PD-(D/E)-XK fold follows a metal-dependent, in-line displacement mechanism. A general base is required to deprotonate the water nucleophile,

Figure 6 DNA-binding by I-SspI. (A) A single I-SspI monomer in complex with a DNA half-site. The regions in direct contact with bases are colored in green. Each distinct contact region on the protein is designated by numbers that correspond to Figure 3 and the text. These regions are magnified to show details in panel B. (B) Schematic diagram of DNA-binding and close-up views of the corresponding contacts. Only half of the DNA target is represented. Residues contacting DNA bases or backbone are labeled as follows: across the DNA, contacts in the minor groove are indicated on the left of each base while contacts in the major groove are indicated on their right. For the protein, residues observed making identical contacts in both monomers (A and B) are labeled in black, whereas residues observed making contacts in individual monomers A and B are labeled in green and blue, respectively. Contacts made by protein to DNA are colored as follows: blue lines indicate direct contacts between bases and protein side chains, blue dashed lines indicate direct contacts between bases and protein main chains, and red lines indicate nonspecific contacts to DNA backbone.

a Lewis acid (usually one or more metal ions) stabilizes the phosphoanion transition state, and an acid protonates the 3'-oxyanion leaving group. The lysine residue in the PD-(D/E)-XK fold is often assigned to the role of a general base, although this role can also be assumed by a variety of other residues. The two acidic side chains of the motif (and occasionally a third acidic side chain) serve to ligate divalent metal ion cofactors, usually Mg^{2+} . Occasionally, an amide-containing Gln or Asn residue can participate in metal binding (Yang *et al*, 2006).

This canonical active site architecture is recapitulated in I-SspI. A single bound calcium ion is observed in each active site in both $|f_o| - |f_c|$ difference maps and in an anomalous difference Fourier map from the native data set collected on a

home X-ray source (Figure 7A). This bound metal ion is coordinated by the scissile phosphate, by a backbone carbonyl oxygen, by Asp 36 from strand $\beta 2$ and by Gln 49 from strand $\beta 3$. An inner shell water molecule bound to this metal ion would be appropriately positioned to act as a nucleophile. Three additional residues (Asp 8 and Gln 11 from helix $\alpha 1$ and Lys 51 from strand $\beta 3$) are also observed in the active site. The lysine residue is located appropriately to participate in general acid-base catalysis and deprotonate the water nucleophile.

It is possible that a second metal ion is also bound by the wild-type enzyme active site, using Asp 8 and/or Glu 11 as coordinating side chains. Modeling a second metal ion near these side chains would recapitulate the structure and

Figure 7 The active site of I-Ssp6803I. (A) The active site of I-Ssp6803I is shown as a ball-and-stick representation. The observed calcium ion position is shown as a red sphere. The anomalous difference map calculated from a native data set collected on a rotating anode X-ray source (CuK α ; $\lambda = 1.54 \text{ \AA}$) is shown in blue and contoured at 4.5σ . The predicted location of the water nucleophile and direction of its attack is indicated by the arrow; the scissile phosphodiester bond is indicated with a red star. (B) Superimposed active sites of I-SspI with *EcoRV* and Hjc. K51, D36 and E11 are conserved; Q49 is replaced by D and E, respectively in *EcoRV* and Hjc.

mechanism observed for a number of restriction endonucleases (Figure 7B).

Fold prediction versus structure determination

At the time that this study was initially submitted for publication, the same structural fold prediction (of a PD-(D/E)-XK domain) was also described by another group, using the same computational server (Orlowski *et al*, 2007). Both predictions produce a reasonable model of the catalytic core region, with an r.m.s.d. across that core of approximately $2 \text{ \AA}$ as compared to the crystal structure. As expected for homology models, more significant differences are observed when they are compared to the actual structure of the entire monomer (Supplementary Figure S4), corresponding to an overall r.m.s.d. between models and structure of $\sim 3.5\text{--}3.7 \text{ \AA}$.

Unfortunately, homology modeling of the DNA-bound I-SspI (Orlowski *et al*, 2007) breaks down significantly, owing to incorrect assignment of the endonuclease quaternary structure. The published attempt to model this complex, using reference models of existing dimeric PD-(D/E)-XK endonucleases (such as *BglI*) lead to a model of I-SspI for subunit packing and for the corresponding DNA conformation which is incorrectly constrained (Supplementary Figure S4). As a result, virtually all of the observed DNA-contacting regions in the I-SspI crystal structure are not predicted in the model of the protein–DNA complex. In particular, the L1 and L1' loops that are associated with the central DNA base pairs of the target site are modeled as extensions of the active site α -helix, and 'region 4' (Figure 6) is not in proximity to DNA.

The homology modeling exercise reported previously (Orlowski *et al*, 2007), and also conducted in the early stages of this study, were very useful for identifying a previously unexpected evolutionary relationship between intron homing and the restriction fold family, and (in our case) for design of crystallizable protein constructs. However, the differences between the model and the crystal structure illustrate the difficulties of predicting both elaborations on core folds and of long-range quaternary interactions and oligomery. Such additional aspects of structure are critical for the functional

and mechanistic constraints placed on any molecular system, including the homing endonuclease described here.

Biological roles and evolutionary successes of nuclease fold families

Enzymatic catalysis of phosphoryl transfer reactions is a fundamental requirement for virtually all forms of nucleic acid modification (Yang *et al*, 2006). A relatively small number of core folds are found to encompass the vast majority of enzymes that make and break phosphodiester bonds. In particular, two unrelated protein folds, the PD-(D/E)-XK and HNH domains, are each found in enzymes involved in similar processes. However, these families enjoy different levels of representation within these processes: the PD-(D/E)-XK family dominates bacterial restriction (but is now shown to have ventured at least once into mobile introns and homing), whereas the HNH family dominates many lineages of mobile introns (and is also found in bacterial colicins) but is found only rarely in bacterial restriction endonucleases.

There are a variety of reasons that might explain the differential success of these protein folds. What is clear is that the PD-(D/E)-XK fold is used frequently, and with great success, to recognize short DNA sequences with absolute fidelity, whereas it is used in at least one limited case to recognize long DNA sequences with reduced fidelity. The comparison of how this fold operates under two separate biological contexts provides an excellent illustration of the balance of mechanistic and structural pressures that dictate the final success and use of such a motif.

Finally, it should be noted that the involvement of the PD-(D/E)-XK fold in the 'dark side' of prokaryotic genetics (as a selfish agent capable of invasion of bacterial genomes) versus its usual and commonly accepted role as a guardian of the bacterial genome (as a restriction endonuclease) may not be as clear-cut as is implied in this study or in conventional views of molecular biology. Restriction–modification (R/M) systems, which were discovered on the basis of their ability to inhibit phage infections of host bacterial strains, appear to be capable of acting as invasive elements on their own

(Kobayashi, 2001). They are often associated with a mobile DNA vectors such as plasmids, viruses and transposons, and may engage in extensive horizontal transfer between bacterial genomes. A key factor in this behavior is the competitive advantage that such an element may exhibit upon incorporation and expression in a host: deletion of the R/M system may lead to death of host progeny as newly replicated (and 'unprotected') DNA target sites are cleaved by residual endonuclease activity. At least one study has demonstrated that the *EcoRI* gene displays homing-type mobility when placed into the appropriate sequence context (Eddy and Gold, 1992). Thus, the use of the restriction nuclease fold in a bacterial homing endonuclease may not represent a sudden departure down a dark path of transposition and non-Mendelian inheritance, but rather a return to one of its fundamental genetic and biological properties.

Materials and methods

Screening for active-site residues

Six mutagenesis primers (Supplementary data) were designed to change nine aspartates (Asp 8, 31, 36, 40, 65, 85, 87, 105 and 120) to alanines. These primers were combined at equimolar concentration in one Stratagene Multi-Change[®] mutagenesis reaction, following the manufacturer's reaction protocol. The reaction products were transformed into *E. coli* BL21 (DE3) RIL⁺ cells and plated on LB containing 50 µg/ml ampicillin.

Fold recognition and comparative modeling

The 3D-Jury (Ginalski *et al*, 2003) consensus method was used for fold recognition at the Meta server (<http://bioinfo.pl/meta/>). We used a K*Sync alignment method (Chivian and Baker, 2006) at Robetta server (Kim *et al*, 2004) (<http://robetta.bakerlab.org/>) to improve alignment between the I-SspI query and 1GEF.

Protein purification

Cultures were induced at 16°C for 18 h. Cells were harvested by centrifugation and lysed using a microfluidizer in 400 mM NaCl, 50 mM Tris, pH 7.5 and 10% glycerol. Cell debris was removed by centrifugation, then forced through a 0.2 µm syringe filter and applied to a heparin affinity column. Protein was concentrated and dialyzed against storage buffer (600 mM NaCl, 50 mM Tris pH 7.5 and 10% glycerol). Size-exclusion chromatography using a Superdex-200 column equilibrated against the same buffer was then performed, and the protein was concentrated to 3.5 mg/ml.

Isothermal titration calorimetry

Studies of DNA binding are described in the Supplementary data.

Crystallization, data collection and structural determination

The DNA oligonucleotides used for cocrystallization were purchased from Integrated DNA Technologies (1 µmol scale, HPLC-purified). The oligos were dissolved in H₂O, and complementary DNA strands were annealed by incubating for 10 min at 95°C

followed by slow cooling. The purified mutant protein was mixed with a three-fold excess of DNA duplex (relative to the concentration of enzyme tetramer) and 10 mM CaCl₂.

The best DNA construct identified for cocrystallization consisted of two strands of sequence: 5'-GAGGCCTTCGGGCTCATAACCCG AAGGA-3' and its complement 5'-TCTCCCTTCGGGTATGAGCCC GAAGGCC-3'. The construct forms a pseudo-palindromic, 27 bp duplex with 2 bp cohesive overhangs on the 5' ends. The sequence of base pairs at positions -11 and -9 (which do not display strong protection by bound endonuclease) were changed to match their counterparts at positions +9 and +11 (which are protected by the bound endonuclease), and thereby promote stable binding. The crystals were grown at 4°C by vapor diffusion against a reservoir containing 15–20% MPD and 100 mM MES buffer at pH 6. Crystals grew in 1 week and were harvested into 100 mM MES (pH 6.0), 25% MPD (2-methyl-2,4 pentanediol), 600 mM NaCl, 20 mM CaCl₂ and 10% glycerol and flash-frozen in liquid nitrogen. These crystals diffracted to 3.3 Å at beamline 5.0.2 of the ALS. The data were processed and scaled using the DENZO/ SCALEPACK program package (Otwinowski and Minor, 1997).

To use the dispersive edge of selenium-substituted methionine (SeMet) for phasing, two residues (Leu 16 and Leu 21) were mutated to methionines. These positions were chosen from regions predicted to be distant from the catalytic active site and the DNA-binding interface, based on the homology model described above (Supplementary Figure S1). SeMet-derivatized I-SspI quadruple-mutant (E11Q/F55K/L16M/L21M) was expressed from the BL21 (DE3) *E. coli* strain under growth and media conditions designed to promote selenomethionine incorporation (Doublié, 1997). The purification, crystallization and data collection protocols were the same as described above, except for the addition of 1% H₂O₂ into the crystal harvesting buffer.

Because of radiation decay, two crystals were used to collect at the peak wavelength and the remote wavelength of selenium, respectively, at beamline 5.0.2 of the ALS. Data sets were processed using the DENZO/SCALEPACK software package (Otwinowski and Minor, 1997). Data statistics are summarized in Table I. CNS (Brunger *et al*, 1998) was used to perform a heavy atoms search. The selenium sites were then input into SHARP (La Fortelle, 1997) for phasing using the MAD method. Model building and initial refinement were performed by COOT (Emsley and Cowtan, 2004). The final model was refined to 3.1 Å resolution against the data set collected at the peak wavelength using REFMAC (Murshudov *et al*, 1997). The stereochemistry of the model was monitored with PROCHECK (Lakowski *et al*, 1993); Table I. A total of 98.2% of the non-glycine residues from the enzyme tetramer are located in the allowed regions of the Ramachandran plot. A native data set collected with a rotating copper anode X-ray source was used to generate an anomalous difference map. Two major peaks were visualized in the map at contour level of 4.5σ and assigned as calcium ions in the final model. All figures were generated with MacPymol (DeLano, 2002).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Accession number: The coordinates of the I-Ssp6803I-DNA complex have been deposited in the protein database (2OST).

References

- Belfort M, Perlman PS (1995a) Mechanisms of intron mobility. *J Biol Chem* **270**: 30237–30240
- Belfort M, Reaban ME, Coetzee T, Dalgaard JZ (1995b) Prokaryotic introns and inteins: a panoply of form and function. *J Bacteriol* **177**: 3897–3903
- Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**: 3379–3388
- Belle A, Landthaler M, Shub D (2002) Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev* **16**: 351–362
- Biniszkiewicz D, Cesnaviciene E, Shub DA (1994) Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J* **13**: 4629–4635
- Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev* **17**: 2875–2888
- Bonocora RP, Shub DA (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol* **39**: 1299–1306
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography &

- NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Bujnicki JM, Radlinska M, Rychlewski L (2001) Polyphyletic evolution of type II restriction enzymes revisited: two independent sources of second-hand folds revealed. *Trends Biochem Sci* **26**: 9–11
- Chevalier B, Sussman D, Otis C, Noel AJ, Turmel M, Lemieux C, Stephens K, Monnat Jr RJ, Stoddard BL (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* **43**: 14015–14026
- Chevalier BS, Monnat Jr RJ, Stoddard BL (2001a) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol* **8**: 312–316
- Chevalier BS, Stoddard BL (2001b) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* **29**: 3757–3774
- Chivian D, Baker D (2006) Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res* **34**: e112
- Deibert M, Grazulis S, Sasnauskas G, Siksnys V, Huber R (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat Struct Biol* **7**: 792–799
- DeLano WL (2002) The PyMol User's Manual. DeLano Scientific, Palo Alto, CA, USA
- Doublet S (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **276**: 523–530
- Eddy SR, Gold L (1992) Artificial mobile DNA element constructed from the *EcoRI* endonuclease gene. *Proc Natl Acad Sci USA* **89**: 1544–1547
- Edgell DR, Belfort M, Shub DA (2000) Barriers to intron promiscuity in bacteria. *J Bacteriol* **182**: 5281–5289
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**: 2126–2132
- Flick KE, Jurica MS, Monnat Jr RJ, Stoddard BL (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **394**: 96–101
- Friedhoff P, Meiss G, Kolmes B, Pieper U, Gimadutdinov O, Urbanke C, Pingoud A (1996) Kinetic analysis of the cleavage of natural and synthetic substrates by the *Serratia* nuclease. *Eur J Biochem* **241**: 572–580
- Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **19**: 1015–1018
- Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305–308
- Gowers DM, Bellamy SR, Halford SE (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms. *Nucleic Acids Res* **32**: 3469–3479
- Heath PJ, Stephens KM, Monnat Jr RJ, Stoddard BL (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nat Struct Biol* **4**: 468–476
- Holm L, Sander C (1996) Alignment of three-dimensional protein structures: network server for database searching. *Methods Enzymol* **266**: 653–662
- Jurica MS, Monnat Jr RJ, Stoddard BL (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol Cell* **2**: 469–476
- Jurica MS, Stoddard BL (1999) Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci* **55**: 1304–1326
- Kelley LA, MacCallum RM, Sternberg MJ (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **200**: 499–520
- Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**: W526–W531
- Kobayashi I (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* **29**: 3742–3756
- Komori K, Sakae S, Daiyasu H, Toh H, Morikawa K, Shinagawa H, Ishino Y (2000) Mutational analysis of the *Pyrococcus furiosus* holliday junction resolvase hjc revealed functionally important residues for dimer formation, junction DNA binding, and cleavage activities. *J Biol Chem* **275**: 40385–40391
- Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res* **27**: 2115–2125
- Kuhlmann UC, Moore GR, James R, Kleanthous C, Hemmings AM (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett* **463**: 1–2
- La Fortelle EDB (1997) Methods in Enzymology. *Macromol Crystallogr* **276**: 472–494
- Lakowski RJ, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**: 283–291
- Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Annu Rev Biochem* **62**: 587–622
- Mueller JE, Brysk M, Loizos N, Belfort M (1993) Homing endonucleases. In *Nucleases*, SM Linn, RSL and RJ Roberts (ed.), pp 111–143. Cold Spring Harbor Laboratory Press
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**: 240–255
- Nesbo CL, Doolittle WF (2003) Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci USA* **100**: 10806–10811
- Nishino T, Komori K, Tsuchiya D, Ishino Y, Morikawa K (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure* **9**: 197–204
- Orlowski J, Boniecki M, Bujnicki JM (2007) I-Ssp68031: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics* **23**: 527–530
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326
- Paquin B, Kathe SD, Nierzwicki-Bauer SA, Shub DA (1997) Origin and evolution of group I introns in cyanobacterial tRNA genes. *J Bacteriol* **179**: 6798–6806
- Reinhold-Hurek B, Shub D (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**: 173–176
- Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol* **342**: 43–56
- Spiegel PC, Chevalier B, Sussman D, Turmel M, Lemieux C, Stoddard BL (2006) The structure of I-CeuI homing endonuclease: evolving asymmetric DNA recognition from a symmetric protein scaffold. *Structure* **14**: 869–880
- Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys* **38**: 49–95
- Vanamee ES, Viadiu H, Kucera R, Dorner L, Picone S, Schildkraut I, Aggarwal AK (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J* **24**: 4198–4208
- Xu M, Kathe S, Goodrich-Blair H, Nierzwicki-Bauer S, Shub D (1990) Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* **250**: 1566–1570
- Yang W, Lee JY, Nowotny M (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* **22**: 5–13