

Recognition of a common rDNA target site in archaea and eukarya by analogous LAGLIDADG and His–Cys box homing endonucleases

Norimichi Nomura¹, Yayoi Nomura¹, Django Sussman², Daniel Klein²
and Barry L. Stoddard^{2,*}

¹Iwata Human Receptor Crystallography Project, ERATO, Japan Science and Technology Agency, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan and ²Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. A3-025 Seattle WA 98109, USA

Received September 21, 2008; Revised October 9, 2008; Accepted October 15, 2008

ABSTRACT

The presence of a homing endonuclease gene (HEG) within a microbial intron or intein empowers the entire element with the ability to invade genomic targets. The persistence of a homing endonuclease lineage depends in part on conservation of its DNA target site. One such rDNA sequence has been invaded both in archaea and in eukarya, by LAGLIDADG and His–Cys box homing endonucleases, respectively. The bases encoded by this target include a universally conserved ribosomal structure, termed helix 69 (H69) in the large ribosomal subunit. This region forms the ‘B2a’ intersubunit bridge to the small ribosomal subunit, contacts bound tRNA in the A- and P-sites, and acts as a trigger for ribosome disassembly through its interactions with ribosome recycling factor. We have determined the DNA-bound structure and specificity profile of an archaeal LAGLIDADG homing endonuclease (I-Vdi141) that recognizes this target site, and compared its specificity with the analogous eukaryal His–Cys box endonuclease I-Ppol. These homodimeric endonuclease scaffolds have arrived at similar specificity profiles across their common biological target and analogous solutions to the problem of accommodating conserved asymmetries within the DNA sequence, but with differences at individual base pairs that are fine-tuned to the sequence conservation of archaeal versus eukaryal ribosomes.

Homing is the transfer of an intervening genetic sequence (either an intron or intein) to a homologous host gene that lacks that same element (1–3). Homing endonucleases,

that are encoded by open reading frames embedded within these sequences, promote their genetic mobility by generating double strand breaks in alleles that lack the intervening sequence. Homologous recombination leads to transfer of the element, using the allele that contains the mobile element as a template. In rare cases, homing endonucleases can also be encoded by free-standing genes (4). In either case, homing endonuclease genes (HEGs) are selfish DNA sequences that are inherited in a dominant, non-Mendelian manner, often with profound consequences (5–7).

Homing endonucleases and their mobile introns can invade host genes encoding rRNA, tRNA and proteins, and are found in virtually all microbial genomes—including phage (8), bacteria (9), archaea (10,11), protista (12), and organellar genomes in fungi and algae (13,14). Mobile introns and homing endonucleases display a periodic life cycle in these host genomes: invasion of a DNA target is followed by vertical transmission with gradual accumulation of mutations that lead to loss of endonuclease activity in individual hosts, followed by eventual loss of the endonuclease reading frame and the intron, and subsequent reinvasion by an active homologue (6,15). In the case of phage and bacteria, this cycle causes the persistence of introns in genomes that are otherwise subject to strong purifying selection and streamlining (16).

Homing endonucleases recognize long DNA sequences (14–40 bp), and tolerate sequence variation at individual base pairs within their targets. This attenuated fidelity enables them to adapt to sequence drift in their target and increases their potential for ectopic transfer to new sites. Comparative genomic studies have demonstrated that (i) homing endonucleases and their surrounding intron/intein sequences are usually found in strongly conserved coding regions within essential (or at least highly beneficial) genes, and that (ii) their ‘specificity profile’ (the relative importance of each base pair identity in the target

*To whom correspondence should be addressed. Tel: +1 206 667 4031; Fax: +1 206 260 9177; Email: bstoddard@fhcrc.org

site for recognition and cleavage by the homing endonuclease) correlates strongly with functional constraints imposed at those same DNA bases by the host (17–19).

There are at least five unique structural families of homing endonucleases, each of which has arrived at the optimal balance of specificity versus fidelity that is most suitable for evolutionary success in their host genomes (3). These include two families (LAGLIDADG and His–Cys Box endonucleases) that are found in eukarya and/or archaea (20,21). Many of these endonucleases are encoded within introns that interrupt highly conserved sequences within ribosomal RNA host genes (rDNA). For example, at least 13 separate positions in the large subunit rDNA of green algae chloroplast genomes are interrupted by introns that contain LAGLIDADG HEGs (22). Similarly, the nucleolar rDNA of protists are also frequently interrupted by group I introns (12,23,24), with many of these intron insertion sites uniformly occupied across species ranging from amoeboid flagellates such as *Naegleria* to myxomycetes (slime molds) such as *Physarum polycephalum*. In contrast with those found in algae, these introns are mostly associated with ‘His–Cys Box’ endonucleases, as typified by the enzyme I-PpoI (25–28).

In addition to rDNA, a variety of microbial genes that are mostly involved in the transmission of genetic information (particularly those encoding tRNA, DNA polymerases, ribonucleotide reductase and thymidylate synthase) have also been subjected to repeated invasions and reinvasions by mobile introns and mobile inteins. In some cases, many introns and inteins (with and without homing endonuclease genes) can be found simultaneously within a single host gene (29). Furthermore, homologous genomic target sites in widely diverse host organisms are often occupied by related mobile elements and their associated endonucleases, which are isoschizomeric enzymes that are obviously diverged from a common ancestor (7,22,30).

In contrast to the many cases described above where individual genes have been invaded repeatedly by mobile intervening sequences and homing endonucleases, only one example has been found of the same target site being independently invaded in separate biological hosts by completely different homing endonuclease lineages. In that case, the precise equivalent of the protist 28S rDNA target site described above (recognized by the His–Cys box endonuclease I-PpoI in a slime mold) has also been invaded in hyperthermophilic archaea by mobile introns and their LAGLIDADG endonucleases (this work). In this study, we describe the DNA-bound crystal structure and cleavage specificity profile of one of these archaeal endonucleases (I-Vdi141I, from *Vulcanisaeta distributa*), and compare its DNA recognition profile against I-PpoI. The functional requirements imposed by invasion of this site by these two homodimeric enzymes have led to strikingly similar balances of specificity and fidelity across their targets, including analogous solutions to the problem of a homodimeric endonuclease accommodating necessary asymmetry in the rDNA sequence. At the same time, these two endonucleases also demonstrate unique DNA sequence preferences for 1 bp position in their respective target sites, which corresponds to a single

RNA base that differs in archaeal versus eukaryal ribosomes.

MATERIALS AND METHODS

Substrate plasmids

The target site for the I-Vdi141I endonuclease is a partially symmetric, 24 bp pseudopalindrome that is cleaved on both strands to liberate two mutually cohesive, 4 base 3′ overhangs. Throughout this article, the individual positions of this target site are numbered from position –12 to +12, with the center of the cleavage site located between base pairs ±1. Seven of the 12 bp of each target half site (±1, 2, 3, 6, 7, 9 and 10) are palindromically conserved between half-sites, while the remaining 4 bp of each half site (±4, 5, 8, 11 and 12) differ between the left- and right-half sites.

The wild-type target DNA was generated by annealing the 25 bp complementary oligonucleotides 5′-CCTGA CTCTCTTAAGGTAGCCAAAA-3′ and 5′-TTTGGCT ACCTTAAGAGAGTCAGGA-3′. The duplex was ligated into a pGEM-T Easy vector (Promega, Madison, WI) to yield ‘pWT’ substrate plasmid. Subsequently, a series of pWT derivatives with target sites mutated at individual base pair positions were prepared by cloning appropriate synthetic oligonucleotide duplexes into pGEM-T Easy. All the cloned inserts were verified by sequencing both strands.

Expression and purification of I-Vdi141I protein

The 23S rRNA gene of the hyperthermophilic archaeon *V. distributa* strain IC-141 contains the Vdi141.L1927 intron that harbors an open reading frame encoding the LAGLIDADG homing endonuclease I-Vdi141I (Nomura, N., unpublished data). The I-Vdi141I gene was PCR amplified from *V. distributa* IC-141 chromosomal DNA using primers ‘Vdi-F’ (5′-GGGGCATATGGTACATG AGGAGGATTTTAAGG) and ‘Vdi-R’ (5′-GGGGGG ATCCTTAAAGACTAGTGCTGACATT). The two primers were designed to place an initiation codon (ATG) and an NdeI site at the 5′ end, and a stop codon and a BamHI site at the 3′ end of the I-Vdi141I coding region. The plasmid pVdi was constructed by ligating the PCR product into NdeI-BamHI-digested pET-15b (Novagen, Darmstadt, Germany). Wild-type enzyme was used in biochemical studies of endonuclease activity and specificity (described below). To increase the number of methionine residues for crystallographic phasing, two amino acid substitutions (I65M and I107M) were introduced using QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). The position of the methionine substitutions was chosen based on a homology model of I-Vdi141I that was derived using the crystal structure of the related endonuclease I-CreI, and corresponded to residues assumed to be uninvolved in protein–DNA contacts or catalytic activity. The insert of the resultant expression plasmid pVdiI65M/I107M-10 was sequenced on both strands to confirm the correct sequence.

Escherichia coli BL21-CodonPlus (DE3)-RIL cells (Stratagene) harboring either wild-type I-Vdi141I or pVdiI65M/I107M-10 was grown in LB supplemented

with ampicillin (50 mg/l) at 37°C to an OD₆₀₀ of 0.6, at which time expression of His₆I-Vdi141I was induced by the addition of IPTG to a final concentration of 0.5 mM. The cells were harvested after an additional 5 h of incubation at 37°C. For expression of selenomethionine-derivatized protein, His₆I-Vdi141I was expressed in M9 minimal media in the *E. coli* strain BL21-CodonPlus (DE3)-RIL adapted for growth with methionine pathway inhibition (31). Cells were grown in minimal media at 37°C to an OD₆₀₀ of 0.6, and the following amino acids were added to inhibit the methionine biosynthetic pathway: 100 mg/l lysine, threonine, and phenylalanine; 60 mg/l selenomethionine; 50 mg/l leucine, isoleucine, and valine. Following a 15-min incubation at 37°C, 0.5 mM IPTG was added to induce expression, and the cultures were grown at 37°C for 10–14 h.

Cells were disrupted by sonication in buffer A (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 10% glycerol). The supernatant was incubated at 80°C for 10 min to denature and precipitate most of the *E. coli* proteins. The heat-treated supernatant was dialyzed overnight against buffer B (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole) and mixed with TALON resin (Clontech, Mountain View, CA) at 4°C for 20 min. The bound protein was eluted with buffer C (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 250 mM imidazole), pooled, dialyzed against buffer A. The dialysate was fractionated on a HiTrap Heparin column (GE Healthcare, Chalfant St. Giles, UK). A linear gradient of 0.2–1 M NaCl was used for elution. The His-tag was removed from the protein using a thrombin cleavage capture kit (Novagen). This was followed by chromatography on a Superdex 75 column (GE Healthcare) equilibrated in buffer A. The purified I-Vdi141I was concentrated to 10 mg/ml using an Amicon Ultra-15 spin filter (Millipore, Billerica, MA), and stored in aliquots at –80°C.

Crystallographic structure determination of I-Vdi141I

Oligonucleotides (5'-CTGACTCTCTTAAGGTAGCCA A-3' and its complementary strand 5'-TTGGCTACCTTA AGAGAGTCAG-3') were purchased from Oligo Etc. (Wilsonville, OR; 1 μmole synthesis scale, HPLC-purified research grade). Complementary DNA strands were annealed by incubating for 5 min at 95°C and then allowed to cool to room temperature, generating a 22 bp, blunt-ended duplex. The selenomethionyl-derivatized protein was diluted to 4.6 mg/ml (0.12 mM) and mixed with a 1.4 M excess of DNA duplex at 50°C in the presence of 2 mM Mg²⁺. Immediately after 5 min of incubation at 50°C, the mixture was used for cocrystallization. A 200-μl reservoir containing 9% ethanol, 100 mM MgCl₂, 100 mM HEPES-NaOH at pH 7.5 was equilibrated against a 2-μl drop containing a 1:1 mixture of the protein/DNA complex and the reservoir solution. The crystals (300 μm × 200 μm × 200 μm; space group *P*₆₁, *a* = *b* = 66.58 Å, *c* = 217.91 Å) were grown at 26°C by vapor diffusion within 2–3 days. They were transferred up to reservoir solution containing 25% ethylene glycol and 1% H₂O₂ (used to completely oxidize selenium atoms and improve their X-ray phasing power) and flash frozen in liquid nitrogen.

Table 1. X-ray data and refinement statistics

Crystallographic data*	
Source	5.0.2 ALS
Maximum resolution (Å)	2.32
Resolution of data set (Å)	50.0–2.32 (2.40–2.32)
Wave length (Å)	0.97911
Space group	<i>P</i> ₆ ₁
Unit cell parameters	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.58, 66.58, 217.91
α , β γ (deg.)	90, 90, 120
Number of reflections measured	90 6351
Number of unique reflections	23 652
<i>R</i> _{merge}	0.068 (0.252)
Completeness (%)	99.4 (100)
Redundancy	22.4 (21.4)
Refinement	
<i>R</i> _{work} (%)	23.1
<i>R</i> _{free} (%)	25.0
Resolution (Å)	45.2–2.50
Number of protein atoms	2610
Number of DNA atoms	898
Number of solvent atoms	122
Number of cations	5
r.m.s. deviations	
Bond length (Å)	0.009
Bond angles (deg.)	1.20
Ramachandran plot	
Favorable (%)	92.7
Allowed (%)	7.3
Generous (%)	0
Unfavorable (%)	0
Mean B value (Å ²)	
Overall	43.8
Protein	46.1
DNA	37.4
Solvent	41.9
Cations	46.3

*Statistics for the peak wavelength is given. Data for the inflection wavelength (0.97945 Å) was very similar. Statistics in parentheses are for the highest resolution shell (Å).

The structure of the I-Vdi141I/DNA complex was solved by the multiple anomalous dispersion (MAD) phasing method using selenomethionyl protein. There are three methionine residues in the protein, in addition to the N-terminal residue which is disordered. Data were collected at the ALS synchrotron beamline 5.0.2 (Lawrence Berkeley National Laboratory, Berkeley, CA). Diffraction data were recorded to 2.3 Å resolution (Table 1). Data were processed and scaled using the program HKL2000 (32). Subsequent data analysis was performed using the program CNS (33). Anomalous difference Patterson maps contained exceptionally strong peaks and cross-peaks corresponding to the expected selenium positions. An initial model was built using the program Coot version 0.1 (34). Ten percent of the data was excluded from the refinement, beginning at the initial stages of model building, for the calculation of the cross-validating free R-factor (35). The model was refined with CNS, and the model geometry was checked with program PROCHECK (36). The final refinement statistics were *R*_{work}/*R*_{free} = 0.231/0.250. The structure has been deposited into the protein data bank (RCSB) with the accession code 3E54.

Endonuclease assays and specificity profile determinations

The substrate plasmids described above were linearized with ScaI, which cuts the vector DNA 1.8 kb from the I-Vdi141I target site. The resulting DNA fragments (ca. 3.0 kb) were extracted with phenol–chloroform and precipitated with ethanol. To monitor the cleavage by wild-type I-Vdi141I, substrate DNA (at 20 nM final concentration) was incubated with the enzyme (at 40 nM concentration) in a total volume of 10 μ l of 10 mM Tris–HCl (pH 7.5 at 25°C), 10 mM MgCl₂, 1 mM dithiothreitol and 50 mM NaCl at 90°C for 10 min.

For comparison, the His–Cys box homing endonuclease I-PpoI was purchased from Promega and used as a reference protein in digests using the same panel of target site variants. I-PpoI cleavage assays were performed with substrate DNA (at 20 nM concentration) and the enzyme (0.25 U; \sim 50 nM concentration) in a total volume of 10 μ l of 25 mM CAPS-CHES (pH 10 at 25°C), 10 mM MgCl₂ and 1 mM dithiothreitol at 37°C for 10 min. During incubation, the assay mixtures were overlaid with 25 μ l mineral oil. The reactions were terminated by addition of 5 μ l of 30 mM EDTA. The reaction products were separated by electrophoresis on 1.5% agarose gels and visualized by ethidium bromide staining. The gels were photographed using a GelDoc2000 digital imaging system (Bio-Rad) and the extent of the reaction was determined using Quantity One software version 3.0 (PDI).

RESULTS

Crystal structure of I-Vdi141I/DNA complex

The 23S rRNA gene from the hyperthermophilic archaea *V. distributa* IC-141 is interrupted by a 586 base pair intron, termed Vdi141.L1927 (GenBank accession number: AB178783). The intron contains an open reading frame (ORF) encoding a polypeptide of 169 amino acids with a molecular mass of 19.6 kDa. Alignment with the amino acid sequences of known archaeal homing endonucleases revealed that the intron-encoded protein contains a single copy of a LAGLIDADG-like sequence motif, spanning residues 12–21 (ILGFIEAEG). The assignment of this protein to the homodimeric LAGLIDADG structural family was confirmed when its structure, bound to its DNA target, was determined using X-ray crystallography as described below.

After producing the recombinant protein in *E. coli*, we confirmed that the ORF product displays sequence-specific DNA cleavage activity; the enzyme was therefore designated as I-Vdi141I. The enzyme's substrate corresponds to a 24 bp, partially symmetric pseudopalindrome, containing the intron insertion site, with sequence 5'-CCTGACTCTC↓TTAA↑GGTAGC CAAA-3' (↓ and ↑ indicate cleavage sites on top and bottom strands, respectively). Like all known LAGLIDADG endonucleases, the enzyme generates mutually cohesive, 4 base 3' overhangs. As described below, the protein displays minimal specificity at the outermost base pair of each DNA half-site (positions \pm 12), and in fact was crystallized in complex with a minimal

22 bp DNA duplex (an analogous DNA construct with all 24 bp did not crystallize).

The molecular weight of functional I-Vdi141I was determined to be \sim 40 kDa as estimated by size exclusion chromatography, also implying that the endonuclease is an active homodimer (data not shown). The genomic intron insertion site and basic DNA recognition and cleavage properties of I-Vdi141I described above are quite similar to an analogous intron and its LAGLIDADG homing endonuclease I-ApeKII from the hyperthermophilic archaean *Aeropyrum pernix* K1 (N. Nomura, unpublished results), indicating that this lineage of mobile introns and endonucleases is widespread throughout a significant archaeal host range.

The I-Vdi141I enzyme was crystallized in complex with a blunt-ended DNA duplex corresponding to the core 22 bp sequence of its target site, and the resulting structure of the protein–DNA complex (Figure 1a) was solved at 2.5 Å resolution. The data processing and refinement statistics are presented in Table 1. The crystals were grown in the presence of magnesium, resulting in a bound product complex in which the DNA is cleaved at positions between bases +2 and +3 on the top strand, and between bases –2 and –3 on the bottom strand. I-Vdi141I consists of two symmetric subunits of mixed α/β topology. The core endonuclease fold ($\alpha\beta\alpha\beta\beta\alpha$) is structurally conserved with previously determined structures of LAGLIDADG enzymes (21), including the use of DNA-contacting β -sheets for sequence-specific recognition of the DNA target half-sites (Figure 1b).

The structure of I-Vdi141I is most similar in size and conformation to that of the homodimeric I-CreI and I-MsoI endonucleases (37), which are also encoded within algal rDNA introns but recognize a different target site. Superposition of the DNA-bound structures of I-Vdi141I and I-CreI (Figure 2a) demonstrates conservation of the relative position and length of all secondary structure elements, with an overall root mean square deviation (RMSD) in alpha-carbon positions of \sim 3.8 Å. The difference in structure between these two enzymes is similar regardless of whether a single subunit or the entire dimeric structure is simultaneously superimposed, indicating that the domain interface and packing of the dimers is approximately equivalent.

As well, the overall bend characteristics of the bound DNA are similar between the structures of I-Vdi141I and I-CreI (Figure 2b and c), consisting of overwinding, and simultaneous unstacking (splaying apart) of the central base step at the center of the target site. This characteristic DNA bending, observed for all LAGLIDADG endonucleases in complex with their DNA targets (38), is particularly pronounced for the I-Vdi141I complex, with a negative distortion of the central base pair roll angle (which is normally 0° for unperturbed B-form DNA) by over 50°. In previous calorimetric studies of LAGLIDADG endonuclease target site recognition, similar protein–DNA complexes were shown to be associated with strongly endothermic binding events (38).

The distribution and number of contacts between protein side chains and DNA bases across the target site are similar to other LAGLIDADG homing endonuclease

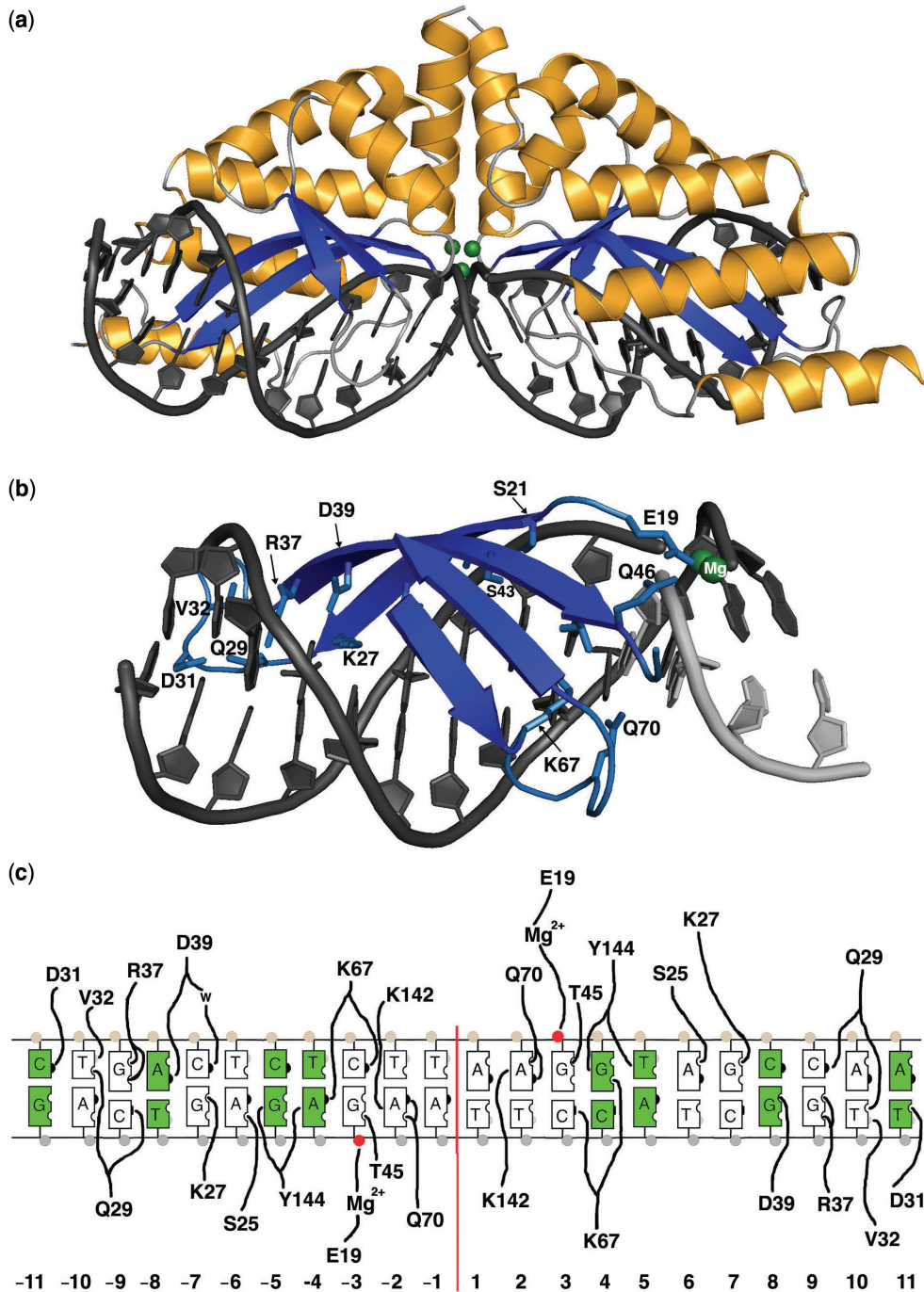


Figure 1. Structure of the I-Vdi14II LAGLIDADG homing endonuclease bound to its cognate DNA target. (a) Ribbon diagram of the homodimeric protein bound to its target site. The DNA is present in a single orientation, providing a unique view of the contacts in each unique DNA half-site. (b and c) Close up and cartoon depiction of the visible contacts between the protein and the DNA. In the schematic, base pairs that are conserved between DNA half-sites are white; those that differ between half-sites are green. Contacts drawn to the right side of each base pair are made in the major groove (to hydrogen bond donors and acceptors that are indicated with bumps and grooves, respectively); contacts drawn to the left side of each base pair are made in the minor groove. The scissile phosphates are shown in red, each contacting a bound divalent cation. Twelve protein side chains are appropriately positioned in each half-site to make contacts to individual nucleotide bases; fifteen out of 36 of the potential hydrogen-bond donors and acceptors in the DNA major groove of each half-site are within contact distance to protein side chains.

complexes (Figure 1c). Approximately 40% (14 of 36) of the potential nucleobase hydrogen bond donors and acceptors in the major groove of each DNA half-site are within contact distance of corresponding protein side

chains; augmented by a smaller number of similar contacts to bases in the minor groove and at least two sequence specific van der Waals contacts. The sequence-specific contacts between DNA and protein are more concentrated

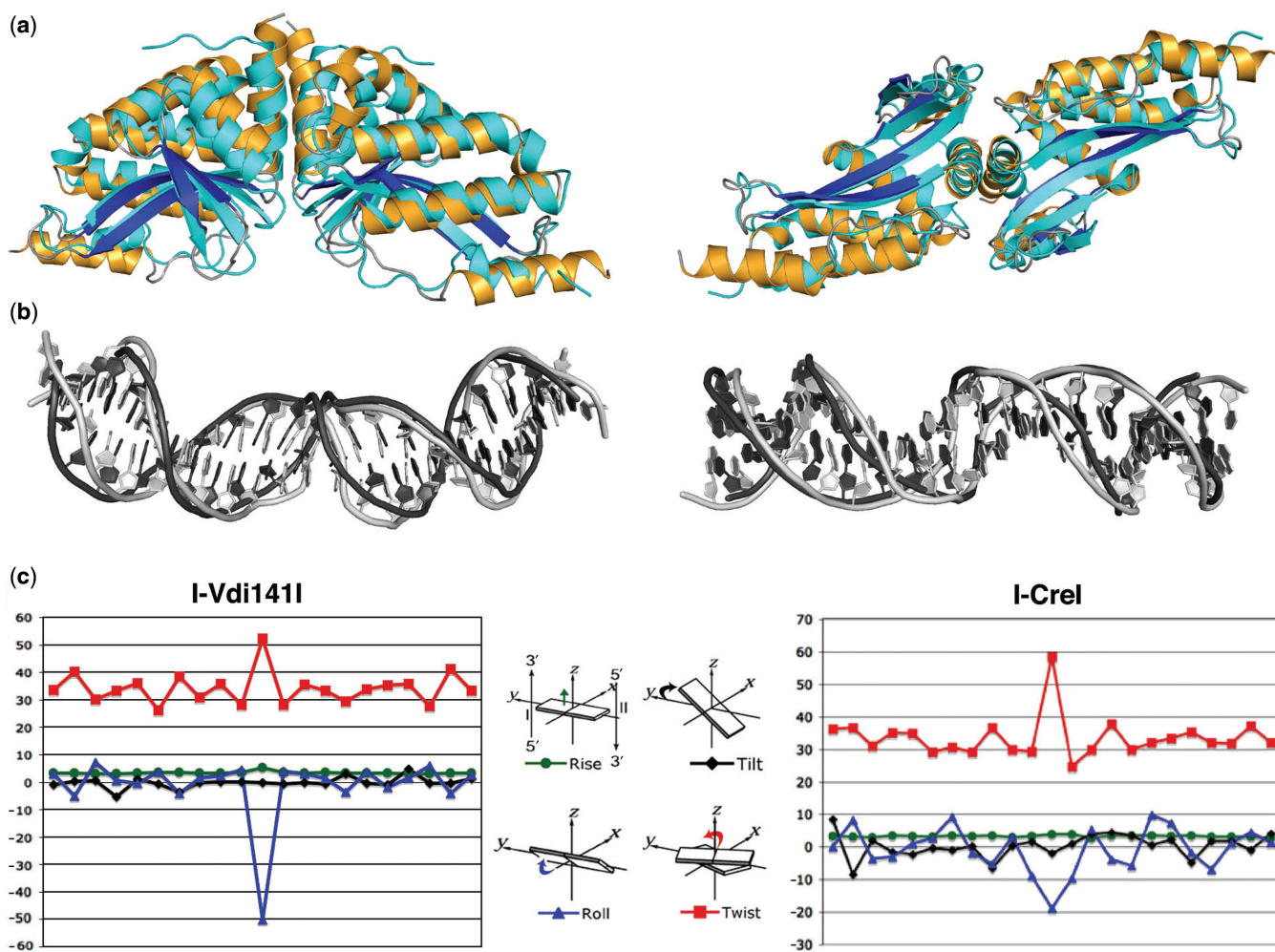


Figure 2. Superposition and comparison of the DNA-bound structures of I-Vdi141I and I-CreI. (a) Superposition of the protein subunits. I-Vdi141I is colored identically to Figure 1; I-CreI is colored cyan. The overall RMSD between all main chain atoms, regardless of whether the superposition is performed using a single subunit or the entire protein homodimer, is ~ 3.8 Å. (b) Superposition of bound DNA duplexes, based solely on superposition of the bound proteins in the two protein–DNA complexes. The DNA target of I-Vdi141I is light grey; the DNA target of I-CreI is dark grey. (c) Conformational parameters of DNA target bound to I-Vdi141I (left) and I-CreI (right). Parameters were quantitated using program ‘Readout’, via a web-based server located at <http://gibk26.bse.kyutech.ac.jp/jouhou/readout/>.

at positions that are conserved between the left and right DNA half-sites (i.e. those base pairs that maintain palindromic symmetry), particularly at positions ± 2 , 3, 9 and 10. The sequence-specific contacts made within the major groove at base pairs ± 2 (by residues Gln 70 and 70') are unique to the LAGLIDADG endonucleases. In all other structures of these protein–DNA complexes, the central four base pair positions (i.e. the nucleotides located between the scissile phosphate groups) are devoid of direct contacts, and are subject only to the influence of sequence-specific DNA conformational preferences that are imposed when the protein binds and bends its target site.

Archaeal I-Vdi141I displays an analogous specificity profile to eukaryal I-PpoI

The rDNA insertion site of the analogous mobile introns in the archaea *Vulcanisaeta* (Vdi141.L1927) and in the

eukaryotic protist *Physarum* (PpLSU3) are almost identical (Figure 3), including exact sequence identity across the core base pairs that are recognized by both corresponding homing endonucleases. Both sites are partially symmetric pseudopalindromes, with symmetry conserved at either seven (I-Vdi141I) or eight (I-PpoI) positions out of twelve base pairs in each DNA half-site. In contrast, these two endonucleases display completely different folds, bound DNA conformations, and contacts across their target sites (Figure 4), while recognizing and cleaving the same substrate sequence.

To compare the specificity of target DNA cleavage by these two enzymes, we performed parallel, systematic cleavage assays with each enzyme, using a series of variant DNA sequences as substrates. Each substrate consisted of the target site of I-Vdi141I, with a single base pair position systematically replaced by one of the three alternative bases. A total of 72 single base-pair variants were therefore prepared, corresponding to 3 different substitutions

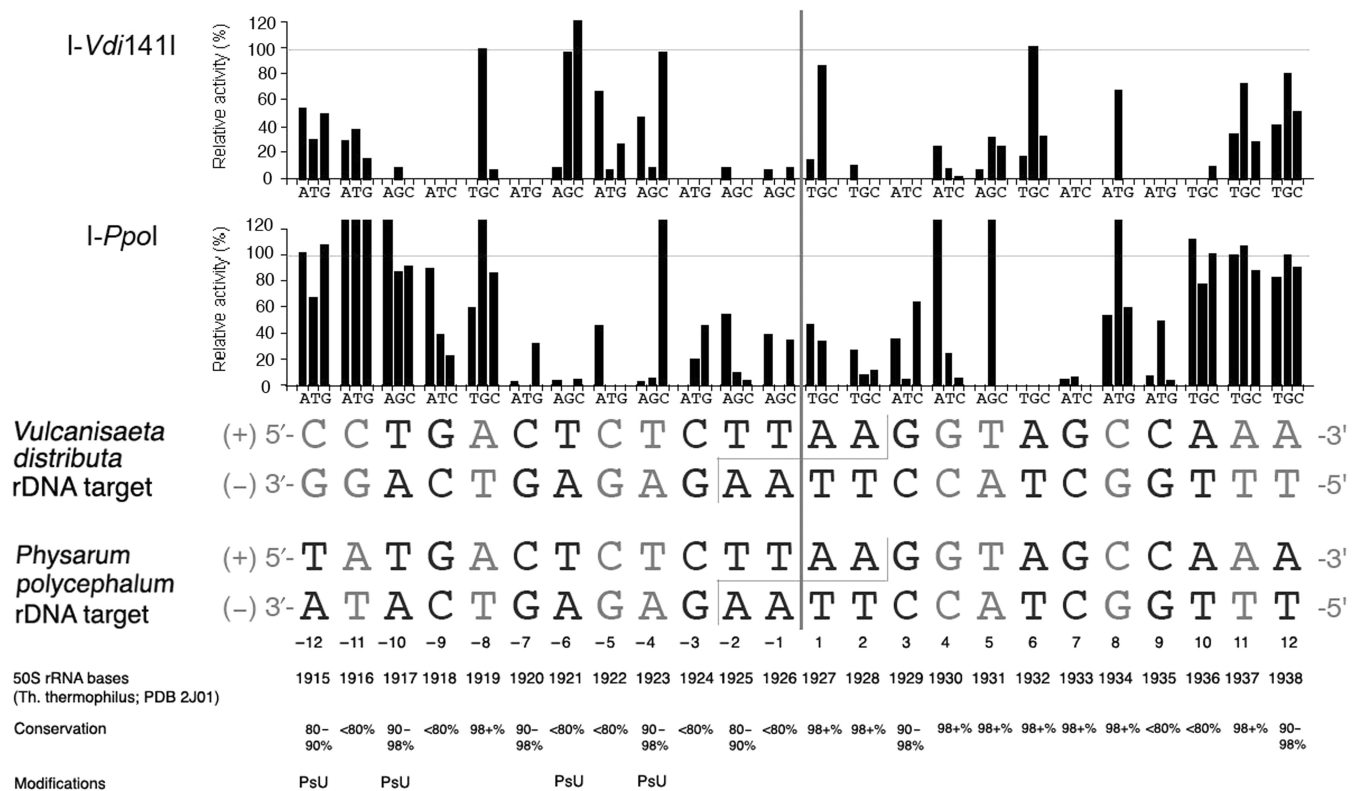


Figure 3. Relative cleavability of individual single base pair variants of the I-Vdi141I target site by I-Vdi141I and I-PpoI. The digest conditions for these experiments are described in Materials and Methods section. The graphs illustrate the relative percent cleavage (Y-axis) of series of linearized DNA target substrates, in which each of three possible basepair substitutions is systematically and individually incorporated at each position across the DNA sequence (X-axis). As described in the Materials and Methods section, digest conditions were chosen such that any substitutions that significantly decrease or increase target site cleavage is visualized. The extent of cleavage of wild-type target site is normalized to 100% in the figure. The two wild-type target sites are identical except for positions -12 and -11, which are outside of the contact and specificity profile of I-PpoI. The red lines denote the center of symmetry of the site and the cleavage pattern of the two enzymes (which both yield 4 base, 3' cohesive overhangs). Shown below the target site sequences are the RNA base numbers of the corresponding rRNA project (taken from PDB structure ID 2J01) and the overall degree of conservation of these bases across all three kingdoms of life (as calculated using the Comparative RNA Website (46) (<http://www.rna.cccb.utexas.edu>)).

at each of 24 separate positions. Each of the variant substrates was incubated with the enzymes for 10 min, and the molar ratio of DNA:protein was adjusted so that no more than 70% of the substrate was converted to the products. The reaction rate was linear as a function of the time over the 10 min time course of the assays. The extents of cleavage observed in these experiments are shown in Figure 3.

Overall, the I-Vdi141I enzyme displays longer target site specificity than I-PpoI, with reductions in relative cleavability observable for one or more base pair substitutions across almost all positions of the entire 24 bp sequence. In contrast, the cleavage specificity of the I-PpoI enzyme extends across a shorter, 18 bp target sequence that is identical to the I-Vdi141I target.

The I-Vdi141I enzyme appears to be more optimized to its physiological rDNA target, with only one substitution in the substrate (-6T to -6C) causing a significant increase in cleavability over the wild-type target. In contrast, I-PpoI is less optimized for recognition of the same target: five substitutions in its 18 bp core target increase cleavability by that enzyme (-8A to -8G, -4T to -4C, +4G to +4A, +5T to +5G and +8C to +8T). In all five of these latter cases, these substitutions increase palindromic symmetry in the target, by converting a base pair

in one DNA half-site to the corresponding sequence of the opposite half-site.

Across the core 18 bp DNA sequence that is recognized with high specificity by both endonucleases, the measured specificity profile is quite similar. Both enzymes are particularly intolerant of base pair substitutions at positions ± 7 , ± 3 and ± 2 , as well as positions -1 and +9 in the left and right half sites, respectively (a total of 8 out of 18 positions). At seven out of the remaining ten positions in the target, the same mutation is 'most tolerated' by both enzymes: -8A to -8G, -6T to -6C, -5C to -5A, -4T to -4C, +4G to +4A, +5T to +5G, and +8C to +8T. Given the completely unique set of protein-DNA contacts and accompanying DNA bend induced by the two enzymes, these similarities in recognition specificity profiles appear to indicate equivalent pressures on both enzyme-intron elements to optimize specificity, relative to the sequences and structural constraints of these rDNA bases in archaea and eukarya.

Finally, two positions in the target sequence, related by symmetry, appear to display significantly different specificities and effects on cleavage for the two enzymes. At position -6, the archaeal I-Vdi141I enzyme slightly prefers a cytosine over the wild type thymine, whereas the

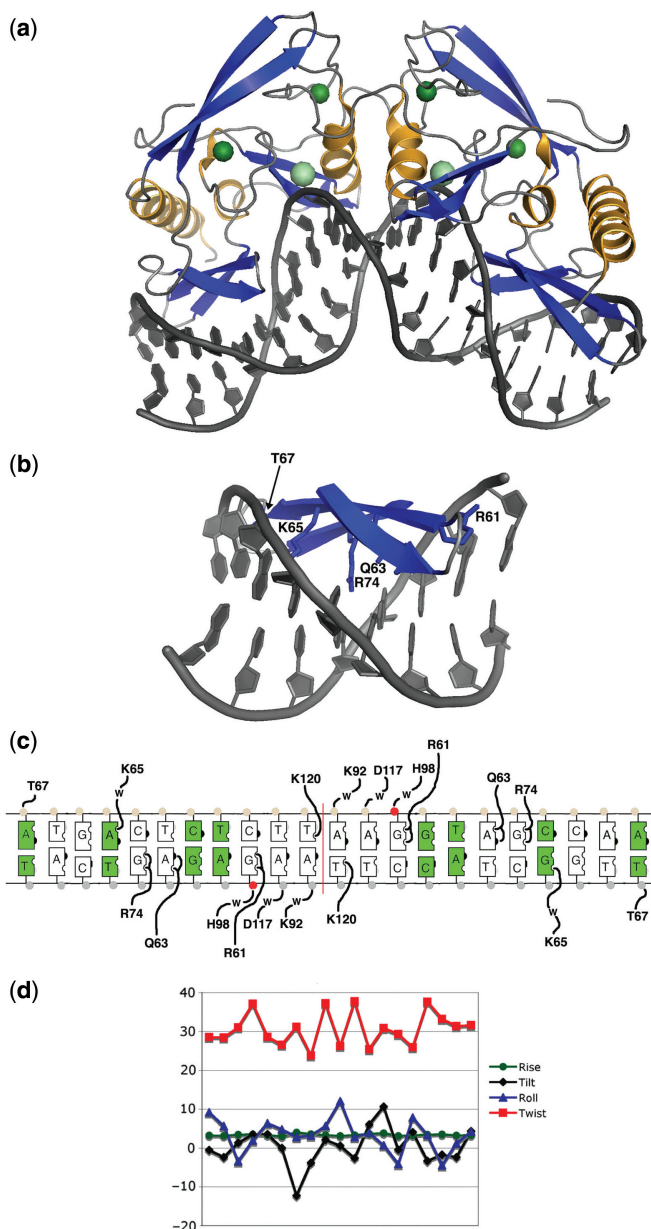


Figure 4. DNA-bound structure, contacts and conformation of I-PpoI. (a) Ribbon diagram of the homodimeric endonuclease bound to its target site. The green spheres are structural zinc ions; the grey spheres are divalent cations (magnesium in this structure) in the enzyme active site. (b and c) Close-up and schematic of contacts between the protein and the DNA target site. Only one half-site is shown in (b) for clarity. The color coding and contacts are shown in the same manner as described in Figure 1. (d) Conformational parameters for I-PpoI bound DNA target, calculated as described in Figure 2.

eukaryal I-PpoI displays strong specificity for the wild type base. In contrast, at position +6 the archaeal enzyme tolerates a substitution from the wild type adenine to a guanine, whereas the eukaryal I-PpoI is again extremely specific for the wild-type base. As discussed below, these variations in cleavage specificities appear to be correlated with subtle differences in the conservation of these bases in archaeal versus eukaryal ribosomes.

DISCUSSION

The rDNA target sequence: a pseudopalindrome fixed by multiple functional and structural constraints

The ribosome bases encoded by the target site for I-Vdi141I (and spanning the shorter, analogous target site for I-PpoI; Figure 3) corresponds to bases 1915–1938 in the large RNA subunit of the intact ribosome (using the numbering system from the structure of the 70S ribosome from *Thermus thermophilus*, PDB entry 2J01). These rRNA bases overlap with the universally conserved ‘helix 69’ (H69) of large ribosomal RNA subunit (bases 1906–1924), a region that is critical for several aspects of ribosome structure and function (Supplementary Figure 1 and 5). The loop of H69 (bases 1912 to 1918) contact helix 44 of the 16S rRNA in the small subunit, forming an inter-subunit bridge (‘B2a’) that is one of only four such RNA structural motifs that are conserved within ribosomes from all kingdoms of life, including those of both mitochondria and chloroplasts (39). These same bases also contact the minor groove of the D-stem junction of tRNA bound in the ribosome’s A-site. Additionally, the stem of helix H69 (bases 1906–1911 and 1919–1924) contacts the D-stem of tRNA bound in the ribosome P-site (39). Four of the bases in this region are subjected to conserved modification via pseudouridylation; elimination of more than two of these modifications in any combination causes significant disruption of ribosome structure and function (40). Binding of ribosome recycling factor (RRF), in combination with elongation factor G, induces a large motion of H69 towards RRF, disrupting bridge B2a and inducing the dissociation of intact ribosomes into subunits for subsequent assembly for the next round of translation (41). Deletion of H69 confers a dominant lethal phenotype, due to defects in ribosome assembly, peptide release and recycling (42).

Of the 22 DNA bases in the I-Vdi141I target site, 13 are conserved across all kingdoms of life at greater than 90% sequence identity; 11 of these positions are located within the shorter target site recognized by I-PpoI (43). Although this DNA target is a pseudopalindrome (and partially overlaps with the H69 stem-loop structure), none of the rRNA bases encoded in the target site are actually base paired with other bases also encoded within the target site (Figure 5a). Thus, the highly conserved, partial palindromic symmetry in this target is a coincidental product of the structural constraints that are imposed by contacts made to ribosomal bases outside the target region. This sequence serves as a high-value target for homing endonuclease cleavage and for occupation by mobile introns, as a result of its very strong sequence conservation and also its accidental symmetry, which facilitates recognition by homodimeric homing endonucleases. These symmetric proteins enjoy an advantage over their larger monomeric cousins, as their shorter coding sequences are more readily tolerated by their host introns with minimal effect on folding and splicing activity.

The ability of symmetric homing endonucleases to display promiscuous base recognition abilities, and thus target asymmetric DNA sequences, is therefore an important feature of both I-Vdi141I and I-PpoI. This capability

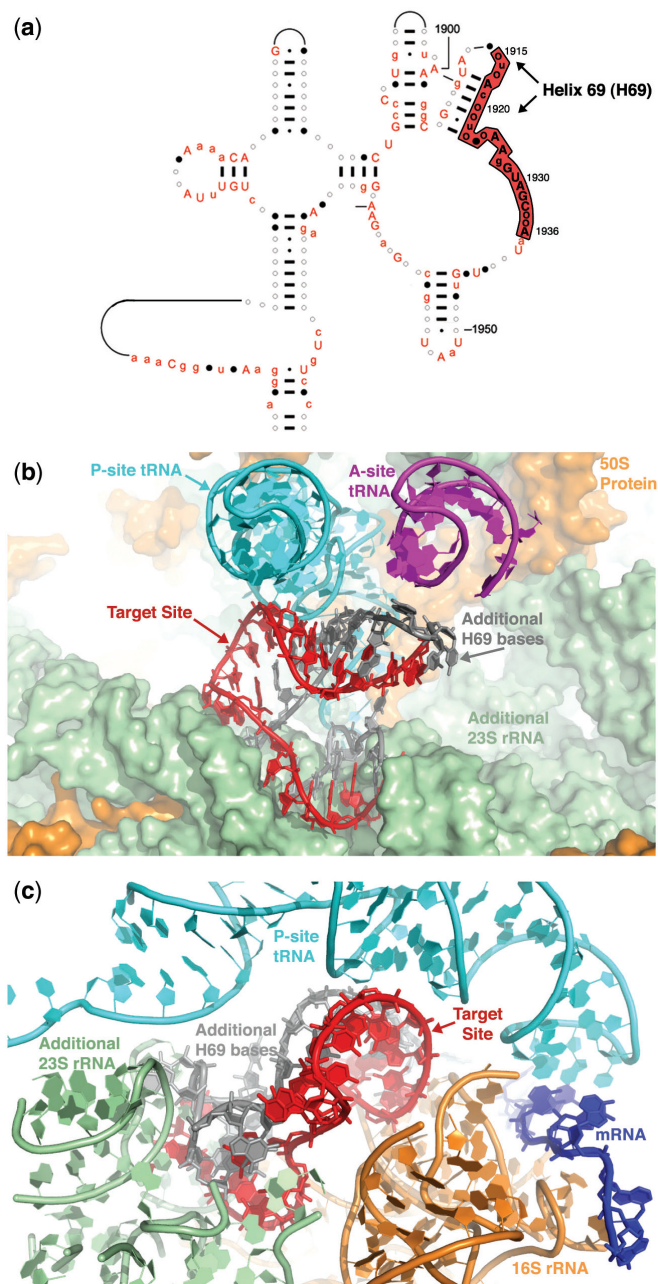


Figure 5. Sequence and structure of the rRNA bases encoded by the target sites of I-Vdi141I and I-PpoI. (a) Local sequence and secondary structure surrounding the target site (outlined in red) and the H69 helix (described in the text; H69 forms a bridge and contacts to the neighboring small RNA ribosome subunit and tRNA molecules bound in the A- and P-sites). The representation of base positions reflects the relative conservation of each position across all three biological kingdoms: open circles represent bases that display <80% identity; closed circles 80–90% identity, lower case bases 90–98% identity, and upper case bases >98% identity. The figure was prepared at the Comparative RNA Website (46) (<http://www.rna.ccb.utexas.edu>) and is used with permission. (b and c) Two separate views of the position of the rRNA bases encoded by the target site (shown in red) in the context of the 70S ribosome structure (PDB ID 2J01). Neighboring subunits and elements (A- and P-site tRNA, 16S and 23S rRNA, mRNA) are labeled and color coded as shown in the figure.

has also been studied in detail for several additional LAGLIDADG homing endonucleases, including I-CreI and I-MsoI [which have been visualized at high resolution (37)] and I-CeuI [another homodimer that displays particularly noteworthy ability to preferentially recognize and cleave a very asymmetric target site (44)]. These studies demonstrate that the ability to accommodate sequence differences at related positions in their target half sites is a function of several features of their protein–DNA contacts: (i) reduced numbers of contacts to potential nucleotide hydrogen bond partners across the target site, (ii) the use of monodentate contacts to DNA bases (such as those made by lysine and serine) rather than bidentate contacts (such as those often formed by arginine, glutamine and asparagine), and (iii) the use of water-mediated contacts (which allows the polarity of hydrogen bond donor-acceptor pairs to be reversed while using the same protein side chain). A comparison of the contact patterns displayed by I-Vdi141I and I-PpoI to their respective DNA targets (Figures 1c and 4c) echo these observations, with a significant number of lysine- and water-mediated contacts made to bases throughout their target sites.

The exceptional ‘quality’ of the rDNA sequence described in this paper as a site for invasion by mobile elements is illustrated even more convincingly by the fact that a region of this same sequence (5′ TAGCCAAA 3′, corresponding to bases +5 to +12 in the endonuclease target) is also the specific site of insertion for the R2 retrotransposon, found in one study to be conserved in the rRNA genes of most insects (45), and probably a considerably wider range of eukarya. As well, the rDNA sequence encoding the ‘domain 4’ region of the large rRNA (the site of the H69 helix) exhibits the highest density of introns in the rRNA encoding genes (46). The high percentage of group I introns in rDNA that are known to be associated with homing endonuclease genes makes it seem likely that this entire region is a hotspot for invasion and persistence. Finally, it also seems likely that the ribosome structure itself, in this region, contributes to the success of invasive introns, by allowing RNA conformations and interactions to form during transcription that facilitate efficient splicing of these intervening sequences.

DNA recognition by the I-PpoI and I-Vdi141I: common specificity profiles produced by disparate mechanisms

As described above, the target sites for these LAGLIDADG and His–Cys box endonucleases are virtually identical, and their specificity profiles are strikingly similar. These recognition events, although closely related phenomenologically, are produced by entirely different mechanisms. There are no analogous contacts made to the same base pair at any position in the two complexes. Whereas I-Vdi141I makes the largest number of direct protein–DNA contacts to base pairs spread across the entire target site (two or more contacts to positions ±2, 3, 7, 9 and 10), I-PpoI concentrates its direct contacts in the middle of each half-site (positions ±3, 6 and 7). As well, the almost identical DNA targets are subjected to completely opposing protein-induced bends. The substrate

of I-Vdi141I (like all previously visualized LAGLIDADG complexes) is subjected to significant unrolling and base unstacking at the exact center of the target site, resulting in a significant narrowing of the minor groove across the four base pairs that are flanked by the scissile phosphates (Figure 2b and c). In contrast, when bound by I-PpoI the same DNA sequence is subjected to significant widening of the minor groove at the site of cleavage, and overall bending of the site approaching 75° (27). This is accomplished with roll angles that never exceed ±12° at any single base pair step, and relatively small distortions of helical winding ('twist') values across the target sites (Figure 4d).

The sites in the DNA target where the two endonucleases most strikingly diverge in their specificity profiles appears to be positions ±6. These positions are palindromically conserved in both targets as an A:T base pair. I-PpoI exhibits maximum specificity at these positions, exhibiting almost complete loss of cleavage activity when any substitution is made in either half-site. In contrast, the archaeal LAGLIDADG endonuclease I-Vdi141I displays significant promiscuity at these same positions, actually exhibiting enhanced cleavage activity at either positions when the A:T base pair is substituted with a G:C. The base pair identity at position -6 (which in the ribosome contributes a base to the stem of the H69 helix) is uniquely conserved in archaea and eukarya, and matches the behavior of the two corresponding enzymes. In archaea, this position is present as a G:C in 73.2% of 41 sequenced archaeal rRNAs (versus 26.8% present as A:T base pairs). In eukarya, this conservation pattern is reversed: the same position in the target site is present at an A:T in 99.4% of 155 sequenced rRNAs (43). Therefore, the I-PpoI appears to be completely optimized for highly specific recognition of the base pair that is found at this position in eukarya. In contrast, the I-Vdi141I enzyme displays lower fidelity at this position (corresponding to the much looser conservation of a single base pair identity in archaea), recognizing both of the potential base pairs that together comprise 100% of the sequenced archaeal ribosome sequences.

Conclusion: unique evolutionary pathways and opposing mechanisms produce analogous genetic and biological outcomes

Since the first structures of DNA-binding proteins in complex with their nucleic acid targets were determined, a major question has been whether there exist any common rules or trends (i.e. a 'code') that influences and dictates recognition specificity. This early hypothesis has yielded to the generally accepted understanding that any such code is best described as set of preferences that are not easily predicted or reproduced in individual DNA-binding proteins. As stated in a recent review, 'It has become clear that there is no simple "code" for protein-DNA recognition and that selecting an optimal binding sequence along the DNA double helix corresponds to more than simply forming a set of specific hydrogen bonds or steric interactions' (47). This was previously clearly demonstrated for homing endonucleases through

a structural comparison of the LAGLIDADG isoschizomers I-CreI and I-MsoI, which display nearly identical protein scaffolds and DNA target sequences, yet only share one-quarter of their DNA-contacting side chains in common.

The comparison of I-Vdi141I and I-PpoI presented here provide a striking demonstration of two completely different protein scaffolds, in radically different organisms and environments, each solving an equivalent and very complex biophysical problem: recognition of a DNA target site with a requisite balance of specificity (to avoid toxicity to the host) and recognition flexibility (to enhance genetic mobility). The convergence of these two proteins to their similar behaviors, albeit with completely different structure-function relationships, illustrates the almost infinite repertoire of functions and behaviors that are endowed upon protein scaffolds through the dual influence of time and evolutionary sequence drift and selection.

Coordinate submission. The structure and X-ray structure factor amplitudes of I-Vdi141I bound to its DNA target site has been submitted to the RCSB database (PDB ID code 3E54).

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENTS

Chromosomal DNA from *V. distributa* IC-141 was provided as a kind gift of Dr. Takashi Itoh. We thank Betty Shen for instruction in model building and refinement; Jill Bolduc and Hitoshi Nakayama for assistance with X-ray crystallography.

FUNDING

The National Institutes of Health (R01 GM49857 and RL1 CA133833 to B.L.S.); the Gates Foundation Grand Challenge Program (to B.L.S.); Grant-in-Aids for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (Nos. 16780232, 16013222 and 14760216 to N.N.). Funding for open access charge: National Institutes of Health and Gates Foundation Grand Challenge Program.

Conflict of interest statement. None declared

REFERENCES

- Belfort, M., Reaban, M.E., Coetzee, T. and Dalgaard, J.Z. (1995) Prokaryotic introns and inteins: a panoply of form and function. *J. Bacter.*, **177**, 3897–3903.
- Dujon, B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene*, **82**, 91–114.
- Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49–95.
- Belle, A., Landthaler, M. and Shub, D.A. (2002) Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes and Dev.*, **16**, 351–362.

5. Cho, Y., Qiu, Y.-L., Kuhlman, P. and Palmer, J.D. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl Acad. Sci. USA*, **95**, 14244–14249.
6. Goddard, M.R. and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl Acad. Sci. USA*, **96**, 13880–13885.
7. Turmel, M., Cote, V., Otis, C., Mercier, J.P., Gray, M.W., Loneragan, K.M. and Lemieux, C. (1995) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). *Mol. Biol. Evol.*, **12**, 533–545.
8. Belfort, M. (1989) Bacteriophage introns: parasites within parasites? *Trends in Genetics*, **5**, 209–213.
9. Bonocora, R. and Shub, D.A. (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA-fMet genes. *Mol. Microbiol.*, **39**, 1299–1306.
10. Dalgaard, J.Z., Garrett, R.A. and Belfort, M. (1993) A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl Acad. Sci. USA*, **90**, 5414–5417.
11. Nomura, N., Morinaga, Y., Kogishi, T., Kim, E., Sako, Y. and Uchido, A. (2002) Heterogeneous yet similar introns reside in identical positions of the rRNA genes in natural isolates of the archaeon *Aeropyrum pernix*. *Gene*, **295**, 43–50.
12. Muscarella, D.E. and Vogt, V.M. (1989) A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell*, **56**, 443–454.
13. Dujon, B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell*, **20**, 185–197.
14. Lemieux, B., Turmel, M. and Lemieux, C. (1988) Unidirectional gene conversions in the chloroplast of *Chlamydomonas* inter-specific hybrids. *Mol. Gen. Genet.*, **212**, 48–55.
15. Burt, A. and Koufopanou, V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Gen. Develop.*, **14**, 609–615.
16. Edgell, D.R., Belfort, M. and Shub, D.A. (2000) Barriers to intron promiscuity in bacteria. *J. Bacteriol.*, **182**, 5281–5289.
17. Edgell, D.R., Stanger, M.J. and Belfort, M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J. Mol. Biol.*, **343**, 1231–1241.
18. Gimble, F.S., Moure, C.M. and Posey, K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J. Mol. Biol.*, **334**, 993–1008.
19. Scalley-Kim, M., McConnell-Smith, A. and Stoddard, B.L. (2007) Coevolution of homing endonuclease specificity and its host target sequence. *J. Mol. Biol.*, **372**, 1305–1319.
20. Galburt, E.A. and Jurica, M.S. (2005) His-Cys Box Homing Endonucleases. In Belfort, M., Wood, D., Derbyshire, V. and Stoddard, B. (eds), *Homing Endonucleases and Inteins*, Vol. 16, Springer-Verlag, London, pp. 85–102.
21. Chevalier, B., Monnat, R.J. Jr. and Stoddard, B.L. (2005) LAGLIDADG Homing Endonucleases. In Belfort, M., Wood, D., Derbyshire, V. and Stoddard, B. (eds), *Homing Endonucleases and Inteins*, Vol. 16, Springer Verlag, Berlin, pp. 34–47.
22. Turmel, M., Gutell, R.R., Mercier, J.-P., Otis, C. and Lemieux, C. (1993) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa: three internal transcribed spacers and 12 group I intron insertion sites. *J. Mol. Biol.*, **232**, 446–467.
23. Johansen, S. and Vogt, V.M. (1994) An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell*, **76**, 725–734.
24. De Jonckheere, J.F. and Brown, S. (1998) Three different group I introns in the nuclear large subunit ribosomal DNA of the amoeboid flagellate *Naegleria*. *Nucleic Acids Res.*, **26**, 456–461.
25. Lowery, R., Hung, L., Knoche, K. and Bandziulis, R. (1992) Properties of I-PpoI: a rare-cutting intron-encoded endonuclease. *Promega Notes*, **38**, 8–12.
26. Whittmayer, P.K. and Raines, R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry*, **35**, 1076–1083.
27. Flick, K.E., Jurica, M.S., Monnat, R.J. Jr. and B.L., (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96–101.
28. Wittmayer, P.K., McKenzie, J.L. and Raines, R.T. (1998) Degenerate DNA recognition by I-PpoI endonuclease. *Gene*, **206**, 11–21.
29. Liu, X.Q., Yang, J. and Meng, Q. (2003) Four inteins and three group II introns encoded in a bacterial ribonucleotide reductase gene. *J. Biol. Chem.*, **278**, 46826–46831.
30. Turmel, M., Mercier, J.-P. and Cote, M.J. (1993) Group I introns interrupt the chloroplast *psaB* and *psbC* and mitochondrial *rrnL* gene in *Chlamydomonas*. *Nucleic Acids Res.*, **21**, 5242–5250.
31. Double, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Meth. Enzymol.*, **276**, 523–530.
32. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.
33. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **54**, 905–921.
34. Emsley, P. and Cowtan, K. (2004) Model-building tools for molecular graphics. *Acta Cryst.*, **D60**, 2126–2132.
35. Brunger, A.T. (1993) Assessment of phase accuracy by cross validation: the free R value. *Methods and applications. Acta Crystallogr. D. Biol. Crystallogr.*, **49**, 24–36.
36. Laskowski, R.J., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystall.*, **26**, 283–291.
37. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R.J. and Stoddard, B.L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-Crel and I-MsoI. *J. Mol. Biol.*, **329**, 253–269.
38. Eastberg, J.H., McConnell, Smith, A., Zhao, L., Ashworth, J., Shen, B.W., Stoddard and B.L., (2007) Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res.*, **35**, 7209–7221.
39. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
40. Liang, X.H., Liu, Q. and Fournier, M.J. (2007) rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell*, **28**, 965–977.
41. Pai, R., Zhang, W., Schuwirth, B., Hirokawa, G., Kaji, H., Kaji, A. and Cate, J. (2008) Structural insights into ribosome recycling factor interactions with the 70S ribosome. *J. Mol. Biol.*, **376**, 1334–1347.
42. Ali, I.K., Lancaster, L., Feinberg, J., Joseph, S. and Noller, H.F. (2006) Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Mol. Cell*, **23**, 865–874.
43. Smith, T.F., Lee, J.C., Gutell, R.R. and Hartman, H. (2008) The origin and evolution of the ribosome. *Biol. Direct*, **3**, 16.
44. Spiegel, P.C., Chevalier, B., Sussman, D., Turmel, M., Lemieux, C. and Stoddard, B.L. (2006) The structure of I-CeuI homing endonuclease: Evolving asymmetric DNA recognition from a symmetric protein scaffold. *Structure*, **14**, 869–880.
45. Jakubczak, J.L., Burke, W.D. and Eickbush, T.H. (1991) Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc. Natl Acad. Sci. USA*, **88**, 3295–3299.
46. Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M. et al. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
47. Lavery, R. (2005) Recognizing DNA. *Q. Rev. Biophys.*, **38**, 339–344.