Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species

Karen L. Posey, Vassiliki Koufopanou¹, Austin Burt¹ and Frederick S. Gimble*¹

Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 W. Holcombe Blvd, Houston, TX 77030, USA and ¹Department of Biological Sciences, Imperial College, Silwood Park, Ascot, Berkshire, SL5 7PY, UK

Received April 29, 2004; Revised July 2, 2004; Accepted July 12, 2004

ABSTRACT

Homing endonuclease genes (HEGs) are mobile DNA elements that are thought to confer no benefit to their host. They encode site-specific DNA endonucleases that perpetuate the element within a species population by homing and disseminate it between species by horizontal transfer. Several yeast species contain the VMA1 HEG that encodes the intein-associated VMA1-derived endonuclease (VDE). The evolutionary state of VDEs from 12 species was assessed by assaying their endonuclease activities. Only two enzymes are active, PI-Zbal from Zygosaccharomyces bailii and PI-Scel from Saccharomyces cariocanus. PI-Zbal cleaves the Z.bailii recognition sequence significantly faster than the Saccharomyces cerevisiae site, which differs at six nucleotide positions. A mutational analysis indicates that PI-Zbal cleaves the S. cerevisiae substrate poorly due to the absence of a contact that is analogous to one made in PI-Scel between Gln-55 and nucleotides +9/+10. PI-Zbal cleaves the Z.bailii substrate primarily due to a single base-pair substitution (A/T+5 → T/A+5). Structural modeling of the PI-Zbal/DNA complex suggests that Arg-331, which is absent in PI-Scel, contacts T/A+5, and the reduced activity observed in a PI-Zbal R331A mutant provides evidence for this interaction. These data illustrate that homing endonucleases evolve altered specificity as they adapt to recognize alternative target sites.

INTRODUCTION

Homing endonuclease genes (HEGs) are mobile DNA elements that occur in numerous fungi, protists, bacteria and viruses (1,2). Similar to transposons, another class of mobile elements, HEGs inhabit a host organism in order to survive, but they do not transit through an extrachromosomal phase as part of their ‘life cycle’. HEGs have evolved mutually beneficial associations with other molecular elements possessing different activities to minimize their impact on host fitness. They are found situated within self-splicing RNA introns that excise the element at the RNA level or as integral components of inteins that are removed at the protein level by an autocatalytic protein splicing reaction. There is little evidence for any positive selection for intron- or intein-encoded HEGs by the host organism, yet excision of these elements from the host genome does not occur at high frequency because they are often inserted within critical regions of genes that are required for host viability where imprecise deletion would produce a non-functional host protein (3–5).

Phylogenetic analyses and biochemical studies indicate that HEGs ensure their persistence in nature by moving at high frequency among species through horizontal transmission (6–10). This lateral transfer between species genomes may be mediated by the HEG-encoded homing endonuclease that is responsible for propagating the element throughout the population of the new resident species. In HEG+/HEG− heterozygous individuals, homing endonucleases initiate a gene conversion event termed ‘homing’ by introducing a double-strand break in the allele of the host gene that lacks the HEG (11,12). DNA repair of the break using the HEG+ allele as template and host-encoded repair proteins results in the super-Mendelian inheritance of the HEG. Phylogenetic and biochemical evidences indicate that over evolutionary time, the absence of selection for HEGs leads to their degeneration through mutation and eventual elimination (7,13,14). Taken as a whole, the data support a model in which HEGs undergo recurrent cycles of (i) horizontal transmission to new genomes, (ii) spread and fixation in the recipient population by homing, (iii) degeneration due to the absence of intact recognition sequences and (iv) eventual loss (7,8). If invasion of a new species occurs prior to complete degradation of the homing endonuclease activity in the old species, the element continues to persist and the cycle begins anew.

An intein located within the VMA1 gene of several species of yeast exhibits many of the features of the proposed HEG life cycle. Autocatalytic protein splicing of the VMA1 intein generates the VMA1-derived endonuclease (VDE) (12,15,16), which is a member of the LAGLIDADG family of homing enzymes (2). A survey of 24 different yeast species revealed that 14 harbor VDEs that have diverged in amino acid sequence (average divergence was 40%) are inserted at

*To whom correspondence should be addressed. Tel: +1 713 677 7605; Fax: +1 713 677 7689; Email: fgimble@ibt.tamhsc.edu

that more than 120 gains and losses of VDEs have occurred in the 24 different yeast species over a period of about 800 million years (8). Whether a vector-based mechanism or other means transferred the VDE genes between yeast species remains unclear. Homing of the *Saccharomyces cerevisiae* VMA1 intein from donor *HEG* to recipient *HEG* alleles occurs at high frequency in heterozygous diploids during meiosis when a double-strand break is created in the *HEG* allele by PI-SceI, the *S. cerevisiae* VDE protein (12). The VMA1 intein behaves as a molecular parasite since it completes homing by using the same yeast DNA repair proteins that normally mediate Spo11p-induced meiotic recombination (17). It cannot be ruled out, however, that VDEs contribute to host fitness since it has been reported that the *S. cerevisiae* protein is involved in the expression of a high-affinity glutathione transporter (18). Evidence that the VMA1 intein degenerates during evolution has come from a detailed biochemical and mutational analysis of a yeast strain (DH1-1A) related to *S. cerevisiae* that contains both HEG* and HEG* alleles (13). Although the DH1-1A protein differs from PI-SceI at only 11 amino acid positions, one of these substitutions significantly reduces the endonuclease activity of the enzyme. The protein splicing activity, unlike the endonuclease activity, does not appear to have degenerated during evolution. Loss of this activity would result in a large polypeptide insertion within an essential protein that would adversely affect host viability (17).

Elucidating the evolutionary history of PI-SceI has been facilitated by having detailed knowledge of the PI-SceI domain structure, the protein–substrate interaction and the active site architecture. PI-SceI consists of two domains, one that harbors the endonuclease catalytic center and the other that contains the active site for protein splicing (19,20). Within the endonuclease domain there are two subdomains that display 2-fold pseudosymmetry. Each of these is structurally related to the monomeric subunits of intron-encoded homing enzymes like I-CreI that are homodimeric and lack a splicing domain (21,22). It is speculated that gene duplication and fusion of a common ancestral endonuclease gene that encoded homodimeric enzymes gave rise to the monomeric proteins like PI-SceI with internal symmetry [for review see (1)]. Near the center of the symmetry axis within the endonuclease domain are two overlapping active sites that are comprised of two conserved aspartic acid residues that coordinate with the essential Mg\(^{2+}\) ion co-factors (20). Critical base-specific and phosphate backbone contacts are made using residues in both domains to a ~31 bp target site that is unique in the 13 million base-pair genome (20,23).

In this work, VDEs from 12 different yeast species were cloned, expressed and purified in order to assess the frequency of degeneration of the VDE activity. All except two of the newly characterized proteins fail to cleave their own DNA substrate, which suggests that HEG degeneration is a common occurrence. Interestingly, the *Zygosaccharomyces bailii* VDE, PI-Zbal, exhibits altered target site specificity since it cleaves its own substrate significantly faster than the *S. cerevisiae* target, which differs at six different base-pair positions. That homing endonucleases can evolve new DNA interactions has been demonstrated previously for the I-CreI and I-MsoI isoschizomers, where phylogenetic and structural analyses revealed that the proteins make different residue contacts to the same nucleotides (24,25). Here, a combined mutational and biochemical strategy was used to define some of the protein and DNA determinants that underlie the different specificities of PI-SceI and PI-Zbal. Taken as a whole, the data reveal the inherent flexibility of the interface between VDEs and their targets that permits these enzymes to migrate efficiently between species.

**MATERIALS AND METHODS**

**Materials**

Synthetic oligonucleotides used for cloning of substrate plasmids, for mutagenesis and for PCR were purchased from Integrated DNA Technologies, Inc. or from MWG Biotech. DNA polymerases, restriction and DNA-modifying enzymes were obtained from New England Biolabs. Cobalt metal affinity resin (TALON) was purchased from CLONTECH, and SP-Sepharose was purchased from Amersham Pharmacia Biotech. All other chemicals were of reagent grade and were obtained from commercial sources.

**Cloning of VDE alleles**

Synthetic oligonucleotide primers complementary to the N- and C-termini of the intein alleles of the different yeast species were used to generate different PCR products from the genomic DNA. These were gel-purified, phosphorylated and ligated into the BsaBI site of pKSS (26). To facilitate cloning of the PCR products, NcoI and BamHI sites were introduced at the N- and C-termini of the open reading frame (ORF), respectively. The DNA sequence was confirmed by sequence analysis. The intein ORFs were subcloned between the NcoI and BamHI sites of pETPI-SceI C-His (NcoI), a derivative of pETPI-SceI C-His (27) containing an NcoI rather than an Ndel site at the position of the initiation codon. Those intein ORFs that contained internal NcoI sites were subcloned directly into the Ndel and BamHI sites of pETPI-SceI C-His.

**Construction of PI-Zbal A55Q and PI-SceI Q55A variants**

PI-SceI Q55A and PI-Zbal A55Q were constructed using duplex oligonucleotide cassette mutagenesis of pETPI-SceI C-His (27) and pETPI-Zbal C-His, respectively. PI-Zbal R331A and PI-Zbal H329A were generated using mutagenic primers by a PCR mutagenesis strategy from pETPI-Zbal C-His. All constructs were verified using automated DNA sequencing.

**Generation of substrates for DNA cleavage and binding assays**

Plasmids containing single VDE recognition sites were constructed by inserting 42 bp duplex cassettes into the EcoRI and HindIII sites of pBluescript as described previously (23). Clones were verified using automated DNA sequencing. A 1837 bp PCR product was generated from these plasmids using two synthetic primers (5'GGCGCACGGGTCGGGCT-
GAAC-3' and 5'-GGGGCGAAAACTCTCAAGGATCT-TACCGC-3') for use in DNA cleavage reactions. DNA cleavage of the PCR substrate yields 782 and 1055 bp products. A 219 bp product used in DNA-binding assays was generated by PCR and 32P-endlabeled as described previously (27).

**Protein purification**

To express the different VDE proteins, the pET plasmid vectors were transformed into BL21 (DE3) cells. The proteins were overexpressed and purified using cobalt-metal affinity and ion-exchange chromatography as described previously for the purification of PI-SceI (27) except that the final storage buffer contained 300 mM KCl. The *Saccharomyces dairienensis* VDE protein was insoluble in the absence of imidazole; therefore, 300 mM imidazole was maintained in its storage buffer. Protein preparations were at least 95% purified with the exception of the *Saccharomyces castellii* protein, which had equimolar amounts of a low-molecular-weight contaminant. The *Saccharomyces exigus* protein appeared as a tight doublet on SDS–PAGE, possibly due to limited proteolysis. The purified proteins were identified as VDEs by western blot analysis using monoclonal anti-histidine antibodies (Sigma).

**Assay of DNA cleavage activity**

To assay DNA cleavage activity, protein variants (100 nM) were incubated at 37°C with various 1.8 kb substrates (7 nM) as described previously (27). The reaction mixtures were analyzed by gel electrophoresis on 1% agarose gels. Gel images of the gels were generated using a Kodak EDAS 290 system and the extents of DNA cleavage were quantified using Quantity One software (Bio-Rad Laboratories, Inc.).

**Assay of DNA-binding activity by electrophoretic mobility shift assay**

Gel-mobility shift assays were preformed as reported previously (27) using a radiolabeled 219 bp DNA duplex containing a single binding site.

**RESULTS**

**Expression and purification of VDEs from different yeast species**

A survey of 24 different species of saccharomycte yeasts (Saccharomyces, Torulaspora, Zygosaccharomyces and Kluyveromyces) revealed the existence of VDEs within 14 species (S.cerevisiae, S.cario-canus, S.exigus, S.castellii, S.dairienensis, S.unisporus, K.polysporus, Z.bailii, Z.bisporus, Z.rouxii, T.pretoriensis, T.globosa, K.latice and C.tropicalis) (8). To investigate the endonuclease activity of these VDEs, their genes were amplified by PCR and cloned into a bacterial expression vector. The presence of the His6-tag was confirmed on each of the purified proteins by western blot analysis using anti-polyhistidine antibodies (data not shown). The different purified proteins are designated as PI-SceI (S.cerevisiae), PI-ScaIP (S.cario-canus), PI-SexIP (S.exigus), PI-ScsIP (S.castellii), PI-SdaIP (S.dairienensis), PI-SunIP (S.unisporus), PI-KpoIP (K.polysporus), PI-ZbaIP (Z.bailii), PI-ZroIP (Z.rouxii), PI-TprIP (T.pretoriensis), PI-TglIP (T.globosa), PI-KlaIP1 (K.latice strain CBS 2896) and PI-KlaIP2 (K.latice strain CBS 683) in keeping with recently defined nomenclature, where the ‘P’ suffix indicates that endonuclaease activity has not been demonstrated previously (29).

**Endonucleolytic activity of VDE proteins from different yeasts on their respective substrates and the *S.cerevisiae* substrate**

According to the model for the ‘life-cycle’ of VDEs, the endonuclease activity, but not the protein splicing activity, is expected to degenerate over evolutionary time once the elements become fixed within a genome (8). To address whether degeneration has occurred, the endonucleolytic activity of the purified proteins was assayed using DNA substrates containing the respective native target sequences. The reactions were performed using Mg2+ as a metal ion cofactor under conditions that had been optimized for PI-SceI. Only two VDE proteins, those from *S.cario-canus* and *Z.bailii*, cleaved their respective DNA substrates after 6 h of incubation at 37°C (Figure 1A). In accordance with nomenclature conventions, these active proteins are designated as PI-ScaIP and PI-ZbaIP, respectively. Surprisingly, although PI-ZbaIP is less similar to PI-SceI than PI-ScaI (61 and 89% identity, respectively), it is more active. To increase the likelihood of observing DNA cleavage activity for the other enzymes, we relaxed the stringency of the cleavage reaction by substituting Mn2+ in the reaction buffer since it has been shown that this ion generally increases the DNA cleavage rate of PI-SceI and allows it to cleave DNA targets that are inactive in Mg2+ (23). Indeed, greater amounts of cleavage product were generated by PI-ScaIP and PI-ZbaIP in Mn2+ but no endonucleolytic activity was observed for any of the other VDEs under either condition (data not shown). Thus, these proteins may never have possessed an endonuclease activity or, more likely, they may have degenerated. Alternatively, the reaction conditions used here may not permit these VDEs to cleave their DNAs.

Evidence that the degeneration of VDEs has occurred comes from aligning the residues in PI-SceI that are critical or important for catalysis with their analogs from the various VDEs (Figure 2). It is likely that a minimum criterion for the activity of any LAGLIDAGD enzyme is the presence of conserved acidic residues that correspond to two essential PI-SceI amino acids (Asp-218 and Asp-326) that coordinate the metal ion co-factors (20,30,31). Given this assumption, it is predicted from Figure 2 that most of the VDEs analyzed here are inactive because one or both acidic residues are absent, while four proteins (PI-ZbaIP, PI-ScaIP, PI-TprIP and PI-TglIP) have the potential to be active. Unlike the acidic residues at the catalytic centers, the neighboring amino acids (Asp-229, Lys-301, Thr-341 and Lys-403 in PI-SceI) are not strictly conserved. Mutagenesis of these residues affects PI-SceI activity (27,32,33), and they likely contribute a key role in organizing an extensive network of water molecules at the active sites that...
is involved in catalysis (20,34). It cannot be predicted how residue substitutions at these positions affect hydration of the solvent pocket. Thus, of the four proteins predicted to be active based on the single criterion for activity, two were shown to be functional in biochemical assays (Figure 1A).

The different VDE proteins were also assayed on the *S.cerevisiae* substrate to assess their DNA cleavage activity. Only PI-ScaI and PI-ZbaI were active using this substrate, and their cleavage activity was lower than PI-SceI (Figure 1B). One observation from Figure 1 is that PI-ScaI, which is significantly less active than PI-SceI or PI-ZbaI, prefers the *S.cerevisiae* target to its own. Alignment of the PI-ScaI sequence with that of PI-SceI reveals that only 11% of the residues differ between the two proteins (data not shown). We hypothesize that the low PI-ScaI activity may be due to the substitution of Lys-336 and Lys-376 of PI-SceI, which have been shown to make important phosphate backbone contacts (20,27), with threonine and serine, respectively. What is most interesting is that PI-ZbaI cleaves the *Z.bailii* DNA target significantly faster than the *S.cerevisiae* target. In contrast, PI-SceI exhibits similar activity toward both substrates (Figure 1C and data not shown). Based on these data, we speculate that PI-ZbaI lacks important contacts used by PI-SceI to bind the *S.cerevisiae* target or that steric clashes or electrostatic repulsions prevent PI-ZbaI from binding to the substrate. Second, we hypothesize that contact(s) present in PI-ZbaI permit it to cleave the *Z.bailii* DNA target. Elucidating why PI-ZbaI cleaves the *S.cerevisiae* target poorly and how it recognizes its own target is the focus of the remainder of this work.

Endonucleolytic activity of PI-SceI on the different yeast substrates

VDEs have adapted for horizontal transfer by targeting to a gene (VMA1, the catalytic subunit of the vacuolar H^+^-ATPase) that is well conserved among yeasts and to a highly conserved recognition sequence located within that gene (8). Moreover, VDE proteins may contact many of the same nucleotides within the recognition sequence since the 11 bp within the PI-SceI recognition sequence identified as being more important for activity (23,35) are conserved in the VDE target sites (8). These important base pairs may be conserved...
because mutations at these positions change the amino acid sequence of the ATPase near its catalytic center. We assayed each of the different yeast VDE target sequences with PI-SceI to examine the effect of the various nucleotide substitutions on DNA cleavage activity by this enzyme (Figure 1C). The nucleotide differences among the targets do not occur at the 11 important positions (Figure 1D), with the exception of those with +9 alterations, and it was predicted that all of the substrates would be active. Indeed, PI-SceI cleaves all of the different VDE targets, with significantly faster cleavage evident for the K.lactis, S.castellii, S.dairenensis targets relative to the S.cerevisiae substrate (Figure 1C). Thus, none of the target sequences has diverged to such a degree that it cannot be cleaved using any homing enzyme. The site recognition flexibility exhibited by PI-SceI is consistent with the hypothesis that a VDE can home to similar target sites in closely related species (8). Previously, it was demonstrated that PI-SceI cleaves a S.cerevisiae substrate containing a G/C base pair at position +20 ~+40% slower than the wild-type substrate (23). This sequence is also identical to the S.cario-

### Table 1. PI-SceI residues involved in catalysis and in binding the DNA target site are listed together with the corresponding residues in PI-ZbaI and Pi-ScaI.

<table>
<thead>
<tr>
<th>PI-SceI residue</th>
<th>Contact type</th>
<th>Nucleotide position</th>
<th>PI-ZbaI residue</th>
<th>PI-ScaI residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys-53</td>
<td>P</td>
<td>+11A/+12G</td>
<td>Lys-53</td>
<td>Lys-53</td>
</tr>
<tr>
<td>His-56</td>
<td>P</td>
<td>+7T/+8T</td>
<td>His-56</td>
<td>His-56</td>
</tr>
<tr>
<td>Arg-57</td>
<td>P</td>
<td>+6C/+7T</td>
<td>Arg-57</td>
<td>Arg-57</td>
</tr>
<tr>
<td>Ser-127</td>
<td>P</td>
<td>+19T/+20T</td>
<td>Ser-128</td>
<td>Ser-127</td>
</tr>
<tr>
<td>Ser-129</td>
<td>P</td>
<td>+20T/+21T</td>
<td>Ser-130</td>
<td>Ser-129</td>
</tr>
<tr>
<td>Ser-169</td>
<td>P</td>
<td>+17A/+18C</td>
<td>Ala-170</td>
<td>Ser-169</td>
</tr>
<tr>
<td>Lys-173</td>
<td>P</td>
<td>+11A/+12G</td>
<td>Lys-172</td>
<td>Lys-173</td>
</tr>
<tr>
<td>Ser-227</td>
<td>P</td>
<td>–2C/–3C</td>
<td>Ser-228</td>
<td>Ser-227</td>
</tr>
<tr>
<td>Ala-261</td>
<td>P</td>
<td>–2C/–1A</td>
<td>Ala-262</td>
<td>Ala-261</td>
</tr>
<tr>
<td>Tyr-267</td>
<td>P</td>
<td>–8T/–7G</td>
<td>Tyr-268</td>
<td>Tyr-267</td>
</tr>
<tr>
<td>Asn-274</td>
<td>P</td>
<td>–11G/–10A</td>
<td>Asn-275</td>
<td>Asn-273</td>
</tr>
<tr>
<td>Gly-275</td>
<td>P</td>
<td>–11G/–10A</td>
<td>Asp-276</td>
<td>Gly-274</td>
</tr>
<tr>
<td>Arg-277</td>
<td>P</td>
<td>–11G/–10A</td>
<td>Arg-278</td>
<td>Arg-276</td>
</tr>
<tr>
<td>Leu-280</td>
<td>P</td>
<td>–9A/–8T</td>
<td>Leu-281</td>
<td>Leu-279</td>
</tr>
<tr>
<td>Asn-281</td>
<td>P</td>
<td>–9A/–8T</td>
<td>Asn-282</td>
<td>Asn-280</td>
</tr>
<tr>
<td>Thr-282</td>
<td>P</td>
<td>–9A/–8T</td>
<td>Tyr-283</td>
<td>Thr-281</td>
</tr>
<tr>
<td>Tyr-328</td>
<td>P</td>
<td>+4G/+5A</td>
<td>His-329</td>
<td>Thr-335</td>
</tr>
<tr>
<td>Lys-336</td>
<td>P</td>
<td>+8T/+9T</td>
<td>Ser-337</td>
<td>Ser-361</td>
</tr>
<tr>
<td>Ser-362</td>
<td>P</td>
<td>+7T/+8T</td>
<td>Ser-362</td>
<td>Ser-363</td>
</tr>
<tr>
<td>Asn-364</td>
<td>P</td>
<td>+6C/+7T</td>
<td>Asn-365</td>
<td>Asn-363</td>
</tr>
<tr>
<td>Lys-376</td>
<td>P</td>
<td>–1T/–2G</td>
<td>Asn-377</td>
<td>Ser-375</td>
</tr>
<tr>
<td>Lys-378</td>
<td>P</td>
<td>–1T/+1G</td>
<td>Arg-379</td>
<td>Lys-377</td>
</tr>
<tr>
<td>Tyr-384</td>
<td>P</td>
<td>+8T/+9T</td>
<td>Tyr-385</td>
<td>Tyr-383</td>
</tr>
<tr>
<td>Gln-55</td>
<td>P, m, m</td>
<td>+7T/+8T, +10G, +9T</td>
<td>Ala-55</td>
<td>Gln-55</td>
</tr>
<tr>
<td>Arg-90</td>
<td>M</td>
<td>+15A</td>
<td>Arg-91</td>
<td>Arg-90</td>
</tr>
<tr>
<td>Arg-94</td>
<td>M</td>
<td>+18G</td>
<td>Arg-95</td>
<td>Arg-94</td>
</tr>
<tr>
<td>His-170</td>
<td>M</td>
<td>+16T</td>
<td>His-171</td>
<td>His-170</td>
</tr>
<tr>
<td>Arg-223</td>
<td>M, M</td>
<td>–7G, –8T</td>
<td>Arg-224</td>
<td>Arg-223</td>
</tr>
<tr>
<td>Glu-366</td>
<td>M, M</td>
<td>+3C, +4C</td>
<td>Glu-367</td>
<td>Glu-365</td>
</tr>
<tr>
<td>His-377</td>
<td>M</td>
<td>+1G</td>
<td>His-378</td>
<td>His-376</td>
</tr>
<tr>
<td>Asp-218</td>
<td>Mg</td>
<td></td>
<td>Asp-219</td>
<td>Asp-218</td>
</tr>
<tr>
<td>Asp-229</td>
<td>Mg</td>
<td></td>
<td>Asp-230</td>
<td>Asp-229</td>
</tr>
<tr>
<td>Lys-301</td>
<td>C</td>
<td></td>
<td>Lys-301</td>
<td>Lys-300</td>
</tr>
<tr>
<td>Asp-326</td>
<td>Mg</td>
<td></td>
<td>Asp-327</td>
<td>Asp-325</td>
</tr>
<tr>
<td>Thr-341</td>
<td>Mg</td>
<td></td>
<td>Thr-342</td>
<td>Thr-340</td>
</tr>
<tr>
<td>Lys-403</td>
<td>C</td>
<td></td>
<td>Lys-404</td>
<td>Lys-402</td>
</tr>
</tbody>
</table>

Residues that differ are shown in boldface. Phosphate contacts are represented with ‘P’, minor groove contacts with ‘m’, major groove contacts with ‘M’, catalytic residues with ‘C’ and metal binding residues with ‘Mg’. The Ala-261 and Gly-275 contacts to DNA are through the main peptide chain, and therefore, should be unaffected by side chain substitutions.
mutant substrates relative to wild-type PI-Scel (Figure 3A) is consistent with the existence of the Gln-55/+9,+10 interaction. Unlike other mutations in PI-Scel and its homing site (27,37), neither the Q55A PI-Scel mutation nor the +9,+10 substrate mutations completely eliminate cleavage activity, suggesting that the Gln-55 contact contributes less binding energy than others.

The absence of Gln-55 in wild-type PI-ZbaI contributes to its low cleavage of the S.cerevisiae substrate since the introduction of a glutamine at this position significantly increases the cleavage rate of the protein [38% (A55Q) versus 13% (wild-type) substrate cleavage after 6 h; Figure 3B]. However, other differences between PI-Scel and PI-ZbaI are also responsible since PI-ZbaI A55Q does not cleave the S.cerevisiae substrate as well as PI-Scel (92% cleaved). The introduction of Gln-55 into PI-ZbaI also increases its activity toward the S.bailii substrate; Figure 4B). Unexpectedly, the Zb+20A substrate was cleaved at a slower rate than the Zbailii substrate (data not shown). Taken together, these data demonstrate that +5T contributes the most toward the activity of the Zbailii substrate of the six base-pair differences, with the −5T, −8C and −2T substitutions playing lesser roles.

Identification of PI-ZbaI amino acid residues that make putative contacts to the T/A5 base pair

To elucidate the role of the T/A5 position within the Z.bailii substrate, we inspected the model of the PI-ZbaI/DNA complex for potential protein contacts to this base pair. Arg-331 is predicted to be ~2.2 Å distant from the adenine base of the T/A5 base pair in the Z.bailii substrate (data not shown), and we speculate that its basic side chain makes base-specific contacts to the DNA. The analogous residue in PI-Scel, Thr-330, is not involved in DNA binding (20). His-329, which is also proximal to the +5 position, is analogous to Tyr-328 in PI-Scel, similar DNA contacts as in PI-Scel (Figure 3B and data not shown).

Definition of nucleotides within the Z.bailii substrate that enable DNA cleavage by PI-ZbaI

We reasoned that one or more of the six base-pair differences between the S.cerevisiae and Z.bailii substrates account for the ability of PI-ZbaI to cleave its own target (Figure 4). A ‘minus-ones/plus-ones’ analysis (38) was used to identify the critical positions by selectively adding or removing the six different base pairs in the S.cerevisiae and Z.bailii targets, respectively (Figure 4C). In our biochemical analysis of the ‘plus-ones’ substrates, insertion of a T/A+5 base pair (Sc+5T) into the S.cerevisiae target contributes more to DNA cleavage by PI-ZbaI than any other replacement and its removal from the Z.bailii substrate dramatically reduces DNA cleavage activity (Figure 4A). This is not the only important nucleotide as base-pair changes at the −8, −5 and −2 positions reproducibly increase activity when present in the S.cerevisiae target [−C/T−8 (Sc−8C), T/A−5 (Sc−5T) and T/A−2 (Sc−2T)] and decrease activity when removed from the Z.bailii substrate [Zb−8T, Zb−5C and Zb−2G]. Thus, our initial experiments using singly substituted mutations suggest that four of the six base pairs contribute positively to the cleavage activity of the Z.bailii target.

Different plus-one base-pair substitutions were combined to assess their contribution toward the total PI-ZbaI activity and to reconstitute the activity of the Z.bailii substrate. A substrate containing both the −5T and +5T mutations, which individually contribute the most activity, is cleaved slower than the Z.bailii target (~54% of the −5T+5T substrate is cleaved after 1 h compared with 79% for the Z.bailii substrate; Figure 4B). When the −8C substitution is combined with the other two mutations, cleavage of the triple mutant substrate is equal to that of the Z.bailii target (Figure 4B). Surprisingly, addition of the last of the four substitutions that individually affect activity increases the level of activity above that of the Z.bailii substrate. This observation raised the possibility that either or both of the +2T or +20G substitutions negatively affect DNA cleavage. Indeed, the +2T substitution has a small negative effect on the cleavage of the Z.bailii substrate (Figure 4B). Unexpectedly, the Zb+20A substrate was cleaved at a slower rate than the Zbailii substrate (data not shown). Taken together, these data demonstrate that +5T contributes the most toward the activity to the Zbailii substrate of the six base-pair differences, with the −5T, −8C and −2T substitutions playing lesser roles.
which makes a phosphate contact between the +4 and +5 positions (27,35).

Mutant PI-ZbaI proteins containing individual alanine substitutions at the Arg-331 and His-329 positions were purified and their DNA cleavage activities were assayed. After 6 h of incubation in the presence of Mg\(^{2+}\), only trace levels of DNA cleavage of a T/A\(^{+5}\)A/T\(^{+5}\) mutant substrate are evident (Figure 5). Similarly low levels of activity are observed when the PI-ZbaI R331A protein is incubated with the wild-type Z.bailii substrate, consistent with the idea that Arg-331 contacts T/A\(^{+5}\). His-329 also appears to play a critical role in the protein–DNA interactions since alanine substitution at this position abolishes DNA cleavage activity (Figure 5). Substitution of Mn\(^{2+}\) for Mg\(^{2+}\) dramatically rescues not only the activity of the Zb\(^{+5}\)A substrate, but also the activities of the R331A and H329A mutant proteins. The ability of Mn\(^{2+}\) to restore function to inactive or partially active DNA substrates and proteins has been observed previously (23,27,39). These data indicate that His-329 and Arg-331 are critical for the cleavage of the Z.bailii substrate.

**DISCUSSION**

In the proposed ‘life cycle’ for HEGs, it is suggested that homing endonuclease activity degenerates over evolutionary time once all HEG\(^{+}\) alleles in a population have been converted into HEG\(^{-}\) alleles due to the absence of selection for function (7). However, if, prior to complete degeneration, homing endonucleases evolve new DNA contacts and lose others to acquire a changed specificity, HEGs have the potential to invade new species that harbor variant target site sequences (7–10). To investigate whether these types of events have occurred during evolution, we purified and biochemically characterized 12 homing endonucleases from a diverse set of yeast species that had previously been shown to harbor analogous VDE homing elements (8). The data support the idea that degeneration occurs at high frequency since we find that most of the yeast homing endonucleases are inactive, which may be due to sequence drift of essential catalytic or DNA-binding residues. It cannot be ruled out, however, that activity was not detected simply because special reaction conditions are required that were not tested. During our survey of yeast VDEs, we discovered that the Z.bailii PI-ZbaI enzyme is significantly less active on the S.cerevisiae target site than on the
Z. bailii site, which differs at six positions. A detailed molecular dissection of this large observed difference in target site specificity between related homing enzymes suggests that different protein/DNA contacts may have evolved in the PI-ZbaI and PI-SceI enzymes.

A myriad of different combinations of mutations could potentially have inactivated the degenerate homing endonuclease during evolution, and we did not attempt to define the specific defects of each enzyme. We determined previously that a degenerate VDE from yeast strain DH1-1A differs from PI-SceI at 11 amino acid positions, and we showed that a single substitution accounted for a 10-fold reduction in activity while other changes exerted little effect (13). Given the larger sequence divergence between PI-SceI and the VDEs characterized here, identifying the residues that abolish activity would be more difficult. Several enzymes, including PI-SexIP, PI-ScsIP, PI-SdalP, PI-SunIP, PI-KpoIP and PI-Zr0P, lack one or both of the conserved acidic residues [Asp-218 and Asp-326 in PI-SceI (20)] that coordinate the essential metal ion co-factors (Figure 2), and it was unexpected that these enzymes are inactive for this reason. The other enzymes may be inactive due to binding defects or for other reasons. Two of the enzymes analyzed in this report, PI-SunIP and PI-KlaP, have been shown previously to be inactive (14). The extensive degeneration of the VDE elements parallels that observed in a survey of the 0 intron HEGs that occur inserted in a group I self-splicing intron within the yeast mitochondrial large subunit (LSU) rRNA gene (7). In this case, it was speculated that two of five enzymes were inactive based on the presence of sequence insertions that disrupted the reading frame. In contrast to the VDE and 0 elements, four intron-encoded LSU homing endonucleases from green algae (I-Crel, I-MsoI, I-PakI and I-Cvul) are all active in biochemical assays (24). Moreover, all 15 members of the enzymes related to I-Crel contain potential catalytic residues, suggesting that these proteins are active as well (24). Why there is less evidence of degeneration of the I-Crel-related HEGs compared with the VDE or 0 elements is unclear, but it may reflect a more recent transition of active I-Crel elements into the different species, or the existence of an undefined positive selection for activity.

The divergence of the PI-ZbaI and PI-SceI target site specificities arose due to variations that permit PI-ZbaI to recognize and cleave the Z. bailii sequence and others that inhibit it from cleaving the S. cerevisiae target. Here, we show that the absence from PI-ZbaI of Gln-55, which contacts base pairs +9 and +10 in PI-SceI, partially accounts for why PI-ZbaI is inactive toward the S. cerevisiae cleavage site. Gln-55 is present in only a minority of the surveyed VDEs, including PI-SceI, PI-ScaI and PI-ScsIP (data not shown). Since PI-ScsIP is unrelated to the other two enzymes in a sequence-based lineage analysis (8), and since there is no apparent evidence of co-evolution of Gln-55 and the A/T to G/C substitution site (data not shown), we speculate that the introduction of the Gln-55 interaction occurred randomly in a limited number of the different VDEs and was probably not present in the common ancestor of these proteins. The new minor groove contact that is made by Gln-55 in PI-SceI may underlie its ability to cleave a wide variety of related VDE recognition sequences. As to why PI-ZbaI cleaves the Z. bailii target when it is inactive against the S. cerevisiae sequence, we provide evidence for an important contact between Arg-331 and T/A +2 that is absent in PI-SceI.

Over one-half of the surveyed VDEs contain an arginine or lysine at the position 331 and have the potential to make a similar contact (data not shown). Many of these are distantly related to PI-ZbaI, suggesting that the putative Arg-331 contact was present in a common VDE ancestor, but was lost by PI-SceI and other enzymes.

This finding that homing endonucleases evolve new contacts while losing others has been well-established in the case of the related intron-encoded homing endonucleases I-Crel and I-MsoI, which display only 33% sequence identity, yet recognize very similar target sites. By comparing the sequences of I-Crel and its homologs, it was predicted that similar contacts mediate subunit assembly, but that new contacts have arisen at the protein–DNA interaction surface (24). Indeed, elucidation of the X-ray crystal structure of the I-MsoI homing endonuclease revealed that the extent of the divergence between the enzymes was even greater than predicted since only one-fifth of the I-Crel residues that contact the DNA substrate are conserved in I-MsoI (25). The conclusion that alternative contacts can be made by homing endonucleases to identical base pairs has also been suggested by the recent demonstration that active chimeric VDEs can be constructed from PI-SceI and the C. tropicalis PI-CtrP enzymes (28), which share little identity. Our results go one step further by showing for the first time that in the process of evolving contacts to one substrate, homing endonucleases can lose the ability to interact with related recognition sites. This finding supports the idea that homing endonuclease sequences diversify at high rates, which may increase the frequency of their invasion into new genomes, and that there is weak selective pressure to maintain identical contacts (24,25). Counterbalancing the evolution of broader recognition that facilitates invasion of related species is the selection for narrower specificity that reduces the frequency of potentially deleterious ectopic cleavage.

Does the endonucleolytic activity of PI-ZbaI imply that homing occurs in Z. bailii yeast? Unlike S. cerevisiae, Z. bailii tolerates low pH and is stress resistant, and it has been frequently associated with food spoilage (40–42). Substantial evidence from genetic and flow cytometry studies indicates that Z. bailii is diploid (41,42). An unusual feature of the yeast is that it produces tetrads with mitotic rather than meiotic spores (41,42). Thus, it would be expected that opportunities for homing would be limited due to the absence of nuclear fusion during sporulation. This suggests that the VDE may have spread through the Z. bailii population by homing when the yeast still included a haploid phase in its life cycle. The high activity of the homing endonucleolytic activity indicates that little degeneration of the enzyme has occurred, either because it confers a selective advantage to the organism or because its entry into Z. bailii is recent.

This study demonstrates that homing endonucleases have evolved new target specificities in nature as they adapt to inhabit new host genomes. There is now considerable interest in more rapidly generating altered specificity enzymes in the laboratory. The ability of homing endonucleases to introduce double-strand breaks at the defined loci in complex genomes has been successfully recruited to stimulate gene targeting in mammalian systems through homologous recombination (43–45). As these methods currently require the pre-existence of the homing endonuclease recognition sequence at or near the
gene of interest, their utility would be expanded if it were possible to generate altered specificity enzymes that cleave at endogenous targets. Recently, a two-hybrid genetic strategy that selects for variants of PI-SceI that binds to target sites with different base substitutions has resulted in proteins with markedly shifted specificities (46). Other genetic methods have screened for variant homing endonucleases that gain the ability to cleave at altered recognition sequences (47). These studies have been complemented by rational design strategies that have engineered chimeric homing enzymes that cleave chimeric target sequences derived from the recognition sequences of two parent enzymes (48,49). Elucidating the sources of altered specificity that arise in nature in the PI-ZbaI and PI-SceI proteins may yield new insights that will aid in further development of these novel reagents in the laboratory.

ACKNOWLEDGEMENTS

We thank Dr Hongye Li for performing the western analysis of the purified VDE proteins. We thank Dr Carmen Moure and Dr Florante Quiocho for valuable discussions. This work was supported by a grant from the National Science Foundation for Dr Florante Quiocho and from the Welch Foundation (BE-1452).

REFERENCES


