

Degeneration of a homing endonuclease and its target sequence in a wild yeast strain

Frederick S. Gimble*

Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 West Holcombe Boulevard, Texas A&M University, Houston, TX 77030, USA

Received June 4, 2001; Revised and Accepted August 28, 2001

DDBJ/EMBL/GenBank accession nos AF389404, AF389405

ABSTRACT

Mobile introns and inteins self-propagate by 'homing', a gene conversion process initiated by site-specific homing endonucleases. The VMA intein, which encodes the PI-SceI endonuclease in *Saccharomyces cerevisiae*, is present in several different yeast strains. Surprisingly, a wild wine yeast (DH1-1A) contains not only the intein⁺ allele, but also an intein-less allele that has not undergone gene conversion. To elucidate how these two alleles co-exist, we characterized the endonuclease encoded by the DH1-1A intein⁺ allele and the target site in the intein⁻ allele. Sequence analysis reveals seven mutations in the 31 bp recognition sequence, none of which occurs at positions that are individually critical for activity. However, binding and cleavage of the sequence by PI-SceI is reduced 10-fold compared to the *S.cerevisiae* target. The PI-SceI analog encoded by the DH1-1A intein⁺ allele contains 11 mutations at residues in the endonuclease and protein splicing domains. None affects protein splicing, but one, a R417Q substitution, accounts for most of the decrease in DNA cleavage and DNA binding activity of the DH1-1A protein. Loss of activity in the DH1-1A endonuclease and target site provides one explanation for co-existence of the intein⁺ and intein⁻ alleles.

INTRODUCTION

Numerous families of mobile genetic elements have been identified and characterized in the genomes of all species examined. These elements have often been described as selfish DNA that confer no benefits to the host. Recently, however, it has become clear that some elements are co-opted for new host functions at later stages of their evolution, prompting some to suggest that a better description of the host–element relationship would reflect a continuous spectrum of interactions ranging from parasitism at one end to mutualism or symbiosis at the other (1). Homing endonuclease genes (HEGs) are mobile genetic elements found inserted into host genes in eukaryotes, bacteria, archaeobacteria and viruses that exist as molecular parasites with no known function, but which occasionally

evolve new functions that benefit their hosts (2–4). HEGs frequently occur as part of introns and inteins that contribute RNA and protein splicing activities that excise the intron/intein and the accompanying HEG to generate a functional host protein. HEGs self-propagate to new individuals of a species by 'homing', a gene conversion process that is initiated by HEG-encoded sequence-specific homing endonucleases (5). When individuals harbor both HEG⁺ and HEG⁻ alleles, the HEG⁺-encoded homing endonuclease cleaves the DNA of the HEG⁻ allele at its recognition sequence. Subsequent repair of this DNA using the homologous HEG⁺ allele as a template results in gene conversion of the HEG⁻ allele and creation of a homozygous HEG⁺ cell. As a consequence, >95% of the progeny inherit the HEG, rather than 50%, as expected from Mendelian inheritance.

Homing endonucleases found in group I introns and inteins are categorized into families based on four conserved sequence motifs, LAGLIDADG, GIY-YIG, H-N-H and the His-Cys box (2,3), but it has been suggested that the latter two comprise a single $\beta\beta\alpha$ Me family of endonucleases (6). The LAGLIDADG family is the largest (more than 150 members) (7), and its members contain either one or two copies of the consensus motif. HEGs encoding LAGLIDADG enzymes are found in group I and archaeal introns, in inteins and as free-standing ORFs within the mitochondrial, chloroplast, nuclear and viral genomes of a wide range of host organisms.

One of the best characterized LAGLIDADG endonucleases is PI-SceI from *Saccharomyces cerevisiae*, which is generated by autocatalytic protein splicing of the VMA intein located within the catalytic subunit of the yeast vacuolar H⁺-ATPase. The dual functions of the enzyme, protein splicing and DNA cleavage, are divided between two domains that are evident from the X-ray crystal structure of the protein (8). The protein splicing domain is structurally related to the GyrA mini-intein from *Mycobacterium xenopi* (9) and to the autoprocessing domain of the *Drosophila* hedgehog protein (10), which catalyzes a protein processing reaction related to protein splicing. Amino acid residues in this domain catalyze protein splicing as well as contact the DNA of the PI-SceI recognition sequence (11–13). The other DNA-binding contacts and the endonucleolytic catalytic center are located in the PI-SceI endonuclease domain (11,13–15), which is structurally homologous to the endonuclease domain of PI-PfuI (16), another LAGLIDADG intein, and to I-CreI (17) and I-DmO1 (18), which are intron-

*Tel: +1 713 677 7605; Fax: +1 713 677 7689; Email: fgimble@ibt.tamu.edu

encoded LAGLIDADG endonucleases that lack protein splicing activity.

The pervasiveness of the yeast VMA intein was assessed by examining several different *Saccharomyces* strains for its presence (5). All of the strains examined contained the intein at the *VMA1* locus (henceforth called the intein⁺ allele), but one wild wine yeast strain (DH1-1A), which is tetraploid, also contained an intein⁻ allele of *VMA1*. The co-existence of these two alleles in one strain was surprising since it would be expected that the intein⁻ allele would be converted to an intein⁺ allele by homing. In our preliminary analysis, we found that the endonuclease encoded by the intein⁺ allele and the target site present in the intein⁻ allele functioned as an enzyme and substrate only under relaxed conditions (Mn²⁺ substituted for Mg²⁺ as cofactor) (5). Here, we identify the molecular basis for the reduced activity of the endonuclease and the recognition sequence and interpret these findings in the light of our current understanding of the PI-*SceI*-DNA complex. Interestingly, the extent and nature of the mutations in the intein⁻ allele target sequence parallel those in a *VMA1* target site from the unrelated yeast *Candida tropicalis*. Our findings provide a molecular picture of a homing endonuclease and its target site in the process of degeneration.

MATERIALS AND METHODS

Cloning and sequencing of the *VMA1* genes from yeast strain DH1-1A

Yeast strain DH1-1A, which is described as a fast brewing yeast for wine making, was a gift from Dr Ira Herskowitz. DNA fragments derived from *VMA1* genes containing or lacking an intein were amplified by PCR from DH1-1A genomic DNA prepared by glass bead lysis. The full-length intein⁻ *VMA1* allele was obtained using primers DH1.top (5'-ATTGTAGCGGCC-GCCAATGGCTGGTGCAATTGAAAACG-3') and DH1.bot (5'-AAGAATGCGGCCGCGTTAATCGGTAGATTCAGC-AAATC-3') that anneal to sequences located at the beginning and end of the *S.cerevisiae* *VMA1* coding frame, respectively, and introduce a *NotI* site on both ends of the PCR product. To amplify the internal intein sequence from the intein⁺ allele and the endonuclease target site from the intein⁻ allele, two primers (5'-CATGTAGCGGCCGCTTCCAGGTGCTTTTGGTTG-TGGTAA-3' and 5'-ATAAGAATGCGGCCGAGTAGTACGCTTCATAATTGGTTCTTT-3') were used that anneal to sequences that encode the *VMA1* exons. All amplifications were performed using the standard reaction conditions optimized for the high fidelity VENT polymerase (New England Biolabs Inc., Beverly, MA). After amplification, the resulting PCR products were digested with *NotI* and cloned into *NotI*-digested pBluescript (Stratagene, San Diego, CA). The nucleotide sequences of the *VMA1* genes and gene fragments were analyzed by DNA sequence analysis. Alignments of the *VMA1* genes were generated using CLUSTAL W 1.81 (19). The GenBank accession numbers for the sequences are AF389404 and AF389405.

Cloning and expression of the DH1-1A homing endonuclease gene

In order to overexpress the DH1-1A homing endonuclease, the DH1-1A intein gene was amplified from our DH1-1A intein⁺

clone by PCR using one primer (5'-GTACATATGTGCTTT-GCCAAGGGT-3') that introduces a *NdeI* site at the initiation codon and a second primer (5'-TTCGGATCCGCGACCCAT-TTTCATGGACGACAACCT-3') that converts residue Asn454 to alanine and introduces a downstream *BamHI* site. Following digestion with *NdeI* and *BamHI*, the DNA fragment was cloned into *NdeI* + *BamHI*-digested pET PI-*SceI* C-His (11), which expresses a 479 amino acid PI-*SceI* derivative that includes a polyhistidine C-terminal extension. The conversion of Asn454 to alanine in this construct prevents elimination of the histidine tag by protein splicing.

Construction of substrate plasmids

Plasmid pBS-PI*Sce36* contains a single *S.cerevisiae* PI-*SceI* recognition site located on a 36 bp insert between the *EcoRI* and *HindIII* sites of pBluescript (Stratagene, San Diego, CA) (20). Plasmids pBSCT-36 and pBSDH-36 are identical to pBS-PI*Sce36* except that the 36 bp *EcoRI*-*HindIII* insert contains a PI-*SceI* target sequence derived from the intein⁻ DH1-1A *VMA1* allele or a predicted target sequence from the *C.tropicalis* *VMA1* allele (21), respectively. To construct pBSCT-36 and pBSDH-36, two synthetic oligonucleotides (5'-AATTCTCTACGTT-GGGTGCAGTGAACGTGGTAACGAGATGGA-3' and 5'-AGCTTCCATCTCGTTACCACGTTCCACCGCACCCAAC-GTAGAG-3' for pBSDH-36 and AATTCTCTATGTTGGTT-GTGGTGAACGTGGTAATGAGATGGA-3' and 5'-AGCT-TCCATCTCATTACCACGTTCCACACAACCAACATAG-AG-3' for pBSCT-36) were annealed and cloned into *EcoRI* + *HindIII*-digested pBluescript.

Purification of inteins

His₆-tagged PI-*SceI*, the DH1-1A endonuclease and the PI-*SceI* variants were expressed and purified from *Escherichia coli* as described previously (11). Protein concentrations were determined by measuring the A₂₈₀ using an extinction coefficient of 5.03 × 10⁴ M⁻¹ cm⁻¹ (11).

Site-directed mutagenesis of the PI-*SceI* gene

All site-directed mutations were introduced with oligonucleotide primers into the coding sequence of the PI-*SceI* gene in plasmid pET PI-*SceI* C-His using an overlapping PCR amplification protocol (22). All mutations and inserted sequences were subsequently confirmed by dideoxy sequence analysis.

Endonuclease assay

The plasmid pBS-PI-*SceI36* (11), which contains a single PI-*SceI* recognition site, was linearized with *XmnI* and used for the endonuclease assay. The linearized substrate (7 nM) was incubated with purified endonuclease (14 nM) in a total volume of 15 μl for various lengths of time at 37°C in cleavage buffer (25 mM Tris-HCl pH 8.5, 100 mM KCl, 2.5 mM β-mercaptoethanol and 50 μg/ml bovine serum albumin) containing 2.5 mM MgCl₂ or MnCl₂. The reactions were terminated by addition of stop buffer (5 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.05% w/v SDS and 2.5% w/v Ficoll) and the reaction products separated by electrophoresis in 1× TBE on a 0.9% agarose gel. The gels were photographed using a Kodak EDAS 290 digital imaging system and the extent of reaction was determined using Quantity One software (Bio-Rad, Hercules, CA).

Gel mobility shift assay

The gel mobility shift assay for determining the binding affinities of the wild-type and mutant endonucleases was performed using a 219 bp duplex DNA fragment containing a single *PI-SceI* recognition site that was prepared by PCR and end-labeled with [³²P]ATP. Purified endonuclease proteins were used in gel mobility shift assays as previously described (11). Similar assays were performed to measure the binding of purified *PI-SceI* to 219 bp substrates containing the *S.cerevisiae*, DH1-1A and *C.tropicalis* recognition sequences.

Protein splicing assay

To test the effect of various site-directed mutations on protein splicing activity, amino acid mutations were introduced into the MYB protein, which is comprised of the yeast *VMA1* intein flanked by maltose-binding protein at the N-terminus and the chitin-binding domain at the C-terminus (23). MYB is expressed from plasmid pMYB129 and does not undergo native protein splicing as a result of an Asn454Ala mutation at the intein C-terminus (23). Plasmid pMYB129 also contains a *Bam*HI site that introduces Ala447Gly/Asn448Ser mutations into the MYB reading frame. Using primers corresponding to the start and end of the *VMA1* intein coding region, DNA fragments were generated by PCR amplification from pET *PI-SceI* C-His templates containing different substitutions in the intein gene. These were digested with *Kpn*I and *Age*I and cloned into *Kpn*I + *Age*I-digested pMYB129. This strategy introduces the desired mutations into the MYB gene and restores the wild-type intein sequence at residues 447, 448 and 454. The MYB fusion proteins were expressed in *E.coli* strain ER2267 as described previously (24). The fusion proteins were separated by SDS-PAGE, followed by Coomassie Blue staining and western blot analysis using antibodies raised against maltose-binding protein and the *VMA* intein (New England Biolabs Inc., Beverly, MA).

RESULTS

Cloning intein⁺ and intein⁻ alleles from strain DH1-1A

The pervasiveness of the *VMA* intein among yeast strains was examined after its initial discovery by screening six different *Saccharomyces* strains (*Saccharomyces oviformis*, strain DH1-1A and two strains each of *Saccharomyces norbensis* and *Saccharomyces diastaticus*) by Southern hybridization analysis using an intein-specific probe (5; data not shown). All strains that were examined contained an intein⁺ *VMA1* allele, which indicates that the mobile element has effectively spread between the strains or was present in a common ancestor. Surprisingly, one strain, DH1-1A, which is described as a fast growing strain for wine making, also harbored an intein⁻ *VMA1* allele (data not shown). Alleles of the *VMA1* gene that lack the intein element are present in other fungi, including *Neurospora crassa* (25), the fission yeast *Schizosaccharomyces pombe* (26) and the filamentous fungus *Ashbya gossypii* (27), but there have been no other reports of strains containing both intein⁺ and intein⁻ alleles. To characterize both the intein⁺ and intein⁻ alleles of DH1-1A in more detail, we attempted to clone the full-length *VMA1* genes using primers corresponding to the termini of the *VMA1* coding region in *S.cerevisiae*. This strategy successfully generated a PCR product corresponding

to the 1.9 kb intein⁻ allele, but for reasons that are unclear, the larger 3.2 kb intein⁺ product was not amplified (data not shown). Additional attempts to isolate the full-length allele using other synthetic primers were unsuccessful (data not shown). However, by using internal primers that anneal to flanking N- and C-terminal extein sequences (see Materials and Methods), a 221 bp PCR product was amplified from the intein⁻ allele that includes the homing endonuclease target site and a 1.6 kb product was generated from the intein⁺ allele that includes the endonuclease ORF (data not shown).

DNA cleavage of the degenerate target sequences from DH1-1A and *C.tropicalis* by *PI-SceI* is slower than that of the *S.cerevisiae* target sequence

We sequenced the entire intein⁻ allele of strain DH1-1A in order to further understand why it has not converted to an intein⁺ allele. Translation of the DH1-1A intein⁻ ORF reveals that it codes for a 617 amino acid protein that is 97% identical to the *S.cerevisiae* H⁺-ATPase subunit (data not shown). The residues that are critical for correct folding or stability of *VMA1p*, including Cys284, Cys539, Tyr343 and Gly250 (28,29), are present in the DH1-1A analog. Likewise, the essential hydrophobic residues present at the catalytic center, Phe452, Tyr532 and Phe538, and the acidic Glu286, which is proposed to participate in ATP hydrolysis (29,30), are also conserved in DH1-1A *VMA1p*. Therefore, it is likely that the intein⁻ allele codes for a functional H⁺-ATPase subunit. Whether the intein⁺ allele also produces a functional H⁺-ATPase protein is uncertain since we were unable to amplify the full-length DNA from this allele using the same primers that were used to generate the intein⁻ PCR product, perhaps due to sequence divergence. However, it is clear that the intein⁺ allele is expressed because the DH1-1A endonuclease is generated in a functional state by protein splicing (5).

A comparison of the 31 bp target sequence derived from the intein⁻ allele of DH1-1A with the predicted *PI-SceI* recognition sequence from *S.cerevisiae* reveals that nearly a quarter of the base pairs (7/31) are different (Fig. 1). Interestingly, all of these substitutions occur at codon 'wobble' positions that do not affect the protein sequence of the encoded vacuolar H⁺-ATPase catalytic subunit. Moreover, none of the substitutions occurs at any of the nine essential base pair positions (20) for *PI-SceI*-mediated cleavage in the predicted *S.cerevisiae* target sequence (Fig. 1). In contrast, the predicted target sequence from the intein⁺ allele from DH1-1A matches the *S.cerevisiae* recognition sequence exactly.

Candida tropicalis is a pathogenic yeast that is only distantly related to *S.cerevisiae*, but it contains a *VMA* intein that is 35.9% identical to the *S.cerevisiae* element that is capable of protein splicing (21). Mutations of the catalytic residues in the *C.tropicalis* homing endonuclease are presumed to inactivate the enzyme, suggesting that homing is impossible in this organism (see Discussion). We compared the predicted target site from the *C.tropicalis* *VMA1* locus to the sequence from *S.cerevisiae* to probe the divergence of the homing sites (Fig. 1). The *C.tropicalis* site contains 7 bp differences compared to the *S.cerevisiae* sequence, and five of these mutations are identical to those present in the DH1-1A intein⁻ target sequence. Furthermore, the base substitutions in both the *Candida* and DH1-1A target sequences do not affect the

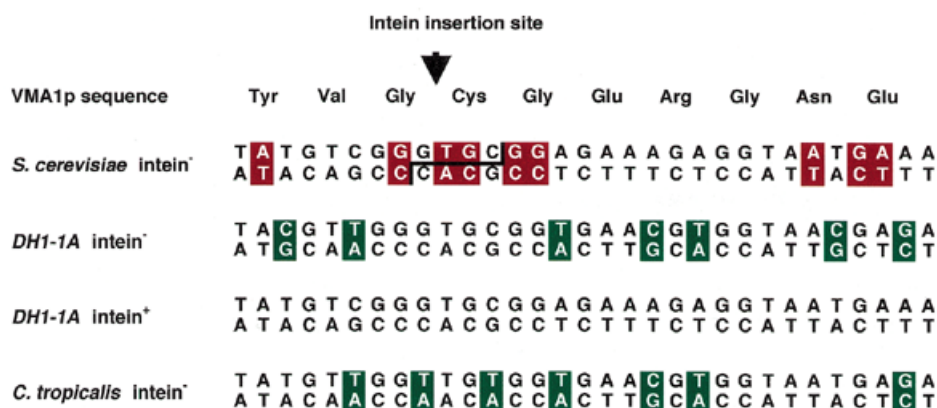


Figure 1. Alignment of the actual and predicted DNA recognition sequences for VMA inteins from *S.cerevisiae*, DH1-1A and *C.tropicalis*. At the top of the figure is shown the protein sequence of the VMA1-encoded H⁺-ATPase in the region overlapping the intein insertion site. The position of the intein insertion site is indicated by an arrow. Base pairs that are shown on a red background in the *S.cerevisiae* intein⁻ sequence are critical for DNA cleavage by PI-SceI (20). The staggered line denotes the site of cleavage by PI-SceI. In the DH1-1A and *C.tropicalis* intein⁻ sequences, the base pairs shown on a green background are those that differ from the *S.cerevisiae* intein⁻ sequence.

Table 1. Cleavage activity of the *S.cerevisiae*, DH1-1A and *C.tropicalis* substrates^a

Target site	Mg ²⁺		Mn ²⁺	
	Activity (×10 ⁻²) ^b	Relative activity ^c	Activity (×10 ⁻²) ^b	Relative activity ^c
<i>S.cerevisiae</i> endonuclease				
<i>S.cerevisiae</i>	2.6 ± 0.3	1	11.5 ± 1	1
DH1-1A	0.28 ± 0.03	0.1	9.4 ± 0.7	0.8
<i>C.tropicalis</i>	0.32 ± 0.03	0.1	7.7 ± 0.2	0.7
DH1-1A endonuclease				
DH1-1A	<0.01	<0.01	14.4 ± 0.2	1.3

^aValues represent means ± SD from at least three experiments.

^bUnits are M DNA cleaved/(min × M enzyme).

^cRelative activity is the ratio of the DH1-1A or *C.tropicalis* to *S.cerevisiae* values.

reading frame of the encoded VMA protein nor do they occur at any of the positions that are essential for cleavage activity.

Previously, we reported that the intein⁻ allele from DH1-1A is resistant to cleavage by wild-type PI-SceI in the presence of magnesium but is cleaved efficiently when manganese is included in the reaction buffer. In those studies, we monitored cleavage of the endogenous recognition site on chromosome IV by Southern analysis using partially purified PI-SceI (5). To test the effect of the DH1-1A and *C.tropicalis* base pair substitutions on cleavage activity in a more defined system, we inserted a single copy of the 31 bp recognition sequences from these strains and from *S.cerevisiae* into pBluescript and assayed substrate cleavage with highly purified protein. Table 1 shows that the rate of cleavage of both the DH1-1A and *C.tropicalis* substrates is reduced 10-fold relative to that of *S.cerevisiae*. Thus, cleavage activity is dramatically affected by the seven mutations within these substrates, even though none of the substitutions occurs at a critical position. Evidently, the cumulative effect of these mutations significantly reduces cleavage activity. Replacement of magnesium

with manganese restores cleavage activity to *S.cerevisiae* substrates that contain single base pair mutations at any of the nine essential positions (20). Table 1 shows that rate enhancements are also observed for the DH1-1A and *C.tropicalis* substrates in Mn²⁺. Although PI-SceI cleaves the DH1-1A and *C.tropicalis* substrates slower than the *S.cerevisiae* substrate in the presence of manganese, it cannot be concluded from the data that this difference is significant.

The DH1-1A and *C.tropicalis* substrates exhibit reduced binding to the PI-SceI endonuclease

We demonstrated previously that most of the *S.cerevisiae* substrates containing single site mutations at critical positions are not cleaved by PI-SceI due to defects in protein binding (20). To test whether the multiple substitutions within the DH1-1A and *C.tropicalis* substrates exert a similar effect, PI-SceI binding to 219 bp DNAs containing single copy inserts of the DH1-1A and *C.tropicalis* recognition sequences was measured by a gel mobility shift assay. Figure 2 shows that as the PI-SceI concentration is increased from 3 × 10⁻¹⁰ to 3 × 10⁻⁹ M, tight binding of the *S.cerevisiae* sequence occurs, but little or no binding to the DH1-1A and *C.tropicalis* sequences is evident. We estimate that the equilibrium binding constants of PI-SceI to the DH1-1A and *C.tropicalis* substrates are ~4 × 10⁻⁹ and >8 × 10⁻⁹ M, respectively (data not shown), which is more than 10-fold greater than that of the *S.cerevisiae* site. Thus, defects in protein binding probably account for the large decreases in DNA cleavage by these substrates. Interestingly, the DH1-1A PI-SceI-DNA complex migrates slightly slower than the *S.cerevisiae* PI-SceI-DNA complex (Fig. 2), which may reflect conformational differences in the DNA of the two complexes (20,31).

Identification of 11 mutations in the homing endonuclease encoded by the DH1-1A intein⁺ allele

Partially purified DH1-1A homing endonuclease only cleaves the chromosomal PI-SceI or DH1-1A target sequences under relaxed conditions in the presence of Mn²⁺ (5). To characterize the DH1-1A endonuclease, the DH1-1A intein gene was amplified by PCR, cloned and sequenced. In addition, we

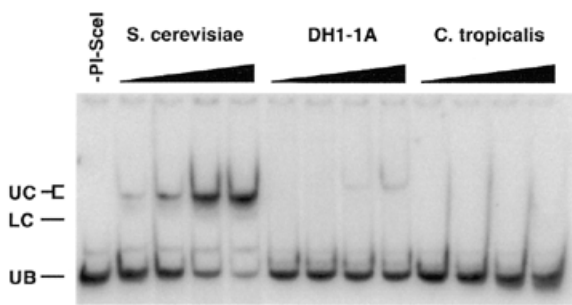


Figure 2. Binding of the predicted endonuclease recognition sequences from *S.cerevisiae*, the DH1-1A intein⁻ allele and *C.tropicalis* to purified PI-SceI protein determined by EMSA. Radiolabeled 219 bp substrates containing one of the different recognition sequences was incubated with increasing concentrations of PI-SceI corresponding to 3×10^{-10} , 7×10^{-10} , 1×10^{-9} and 3×10^{-9} M in each set of four lanes. The samples were separated by electrophoresis in a 7% polyacrylamide gel. The migration positions of unbound 219 bp fragment DNA and the lower and upper PI-SceI-DNA complexes are indicated by UB, LC and UC, respectively.

inserted the gene into an expression vector to generate His₆-tagged protein for purification. Sequence analysis reveals that the DH1-1A homing endonuclease contains 11 different amino acid residues compared to PI-SceI. The DH1-1A endonuclease is 97.5% identical overall to *S.cerevisiae* PI-SceI (data not shown), which is markedly higher than the level of identity between the *C.tropicalis* and *S.cerevisiae* inteins (35.9% identity). Figure 3 shows the positions of the residues in the structure of *S.cerevisiae* PI-SceI that contain amino acid substitutions in the DH1-1A endonuclease. Five of these substitutions (R44S, V67M, I132V, R417Q and D429N) are located within the protein splicing domain. In PI-SceI, Arg44 is located in β 5 and forms a salt bridge on the protein surface with Asp429 in the adjacent anti-parallel β 25. Val67 is part of a flexible loop that contains residues which are thought to contact the phosphate backbone (K.Posey and F.S.Gimble, unpublished data). Ile132 is located within an extended subdomain in a disordered region of the protein and Arg417 is part of β 23. Three substitutions, I276V, V291I and V361A, occur at amino acid residues in the endonuclease domain. Of these, Ile276 is immediately adjacent to Arg277, which is critical for substrate binding, within a disordered region of the apoenzyme crystal structure (32). Val291 and Val361 are part of α 6 and β 21, respectively. The remaining three substitutions, A409V, A411T and A413V, occur at residues that are part of a linker that connects the two domains.

The R417Q substitution in the DH1-1A homing endonuclease reduces its DNA cleavage and DNA binding activities

We purified the recombinant DH1-1A homing endonuclease from *E.coli* in order to characterize its DNA cleavage and DNA binding activities. The His₆-tagged protein was purified using the same procedures as for wild-type *S.cerevisiae* PI-SceI (data not shown). In DNA cleavage assays of the *S.cerevisiae* target sequence, the purified enzyme exhibited 10-fold reduced DNA cleavage activity relative to the *S.cerevisiae* enzyme (Table 2). This finding is consistent with our previous demonstration that the protein does not cleave a genomic DNA

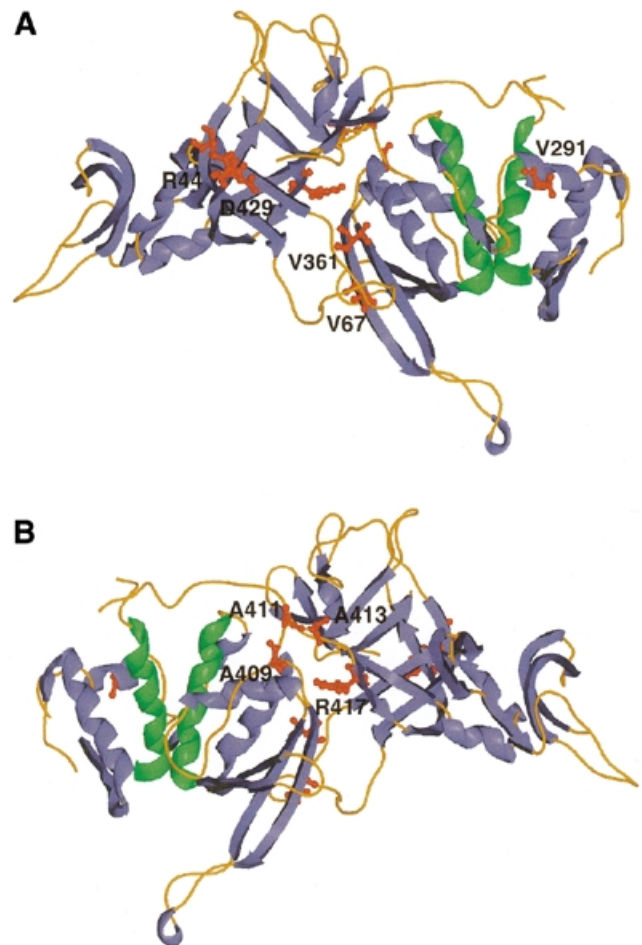


Figure 3. Positions of DH1-1A amino acid substitutions within the PI-SceI crystal structure. (A) A ribbon diagram of the PI-SceI structure (colored purple) was generated using SPOCK (52) from coordinates deposited in the RCSB Protein Data Bank (entry 1DFA). The side chains of PI-SceI residues that are substituted in the DH1-1A protein are colored red and are labeled. The endonucleolytic active site is located at the base of the two conserved LAGLI-DADG α -helices, which are colored green. (B) The same diagram as in (A) but rotated 180° about the y-axis. Not shown are the side chains for Ile132 and Ile276, which are situated in disordered regions of the crystal structure.

substrate. When the assay included Mn²⁺ rather than Mg²⁺, the level of activity was approximately equal to that of wild-type PI-SceI (data not shown). Only trace levels of cleavage activity were evident when the DH1-1A intein⁻ target sequence was mixed with the DH1-1A enzyme in Mg²⁺, which is consistent with an additive negative effect of the DH1-1A endonuclease and DH1-1A substrate mutations (Table 1). Substitution of Mn²⁺ for Mg²⁺ restores the cleavage activity to a level like that of wild-type PI-SceI.

PI-SceI mutants that do not cleave DNA due to defects in DNA binding have their activity restored when Mn²⁺ is substituted for Mg²⁺ in the cleavage reaction buffer (11,13,32). Given the similar behavior of the DH1-1A endonuclease, we hypothesized that its amino acid substitutions reduce DNA binding activity of the protein. Indeed, binding of the DH1-1A protein is ~10-fold reduced compared to *S.cerevisiae* PI-SceI

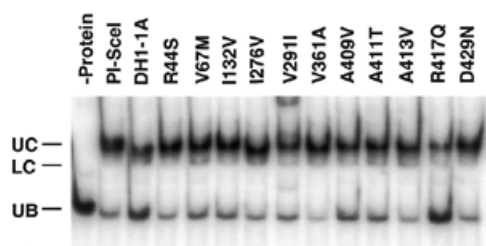


Figure 4. EMSA of binding of PI-*SceI*, the DH1-1A endonuclease and PI-*SceI* variants to a *S.cerevisiae* recognition sequence. The DNA binding activities of PI-*SceI*, of the DH1-1A protein and of 11 PI-*SceI* derivatives, each containing one of the DH1-1A substitutions (amino acid substitutions are indicated above each lane) were determined. The endonuclease protein concentration in each lane is 0.7 nM, which corresponds to the K_d for wild-type PI-*SceI*, and the samples were separated by electrophoresis in a 7% polyacrylamide gel. The migration positions of unbound 219 bp fragment DNA and the lower and upper PI-*SceI*-DNA complexes are indicated by UB, LC and UC, respectively. Similar relative binding affinities were observed using total protein concentrations of 4×10^{-9} and 2×10^{-9} M (data not shown).

in gel mobility shift assays (Fig. 4 and data not shown). In order to determine whether the binding decrease derives from the cumulative effect of all 11 mutations or only from a subset, we inserted the 11 mutations individually into PI-*SceI*, purified the mutant proteins and characterized their DNA cleavage and DNA binding properties. Table 2 shows that all of the proteins but one exhibit DNA cleavage activities that range from 70 to 140% of the wild-type activity. In contrast, the variant that contains the R417Q substitution exhibits 5-fold less activity than wild-type PI-*SceI*. The results of gel mobility shift assays parallel those of the DNA cleavage analysis. Figure 4 shows that the DNA binding activity of the R417Q protein is markedly reduced compared to wild-type PI-*SceI*. In separate DNA binding experiments, we determined that the K_d of the R417Q protein is $\sim 3 \times 10^{-9}$ M (data not shown), which is over 4-fold higher than that of wild-type PI-*SceI*. The binding activities of the other singly substituted variants are more similar to wild-type PI-*SceI* in Figure 4 and in experiments performed at two other protein concentrations (data not shown). Although we cannot conclude that the R417Q mutation is solely responsible for the decrease in cleavage activity of the DH1-1A endonuclease, it must account for a significant portion.

The electrostatic interaction between Arg44 and Asp429 is not required for protein splicing activity

Two of the residues that are substituted in the DH1-1A endonuclease, Arg44 and Asp429, normally form a salt bridge in the PI-*SceI* structure that links anti-parallel β -strands $\beta 5$ and $\beta 25$, which comprise a portion of the cavity wall that bounds the protein splicing catalytic center (Fig. 3). In DH1-1A, Arg44 and Asp429 are replaced by Ser44 and Asn429, which have the potential to form a hydrogen bond in the folded structure. We wondered whether the salt bridge in PI-*SceI* normally plays a critical role in stabilizing the conformation of the protein splicing active site and whether complementary mutations had occurred at residues 44 and 429 in the DH1-1A protein to maintain this interaction. To test this idea, we introduced several mutations at residues 44 and 429 within a VMA intein flanked at the N-terminus by maltose-binding protein and at the C-terminus by the chitin-binding domain and

Table 2. Cleavage activity of *S.cerevisiae* PI-*SceI*, the DH1-1A homing endonuclease and singly substituted PI-*SceI* endonucleases^a

Protein	Activity ($\times 10^{-2}$) ^b	Relative activity ^c
<i>S.cerevisiae</i>	2.6 ± 0.3	1
DH1-1A	0.25 ± 0.05	0.1
R44S	3.0 ± 0.3	1.2
V67M	2.4 ± 0.3	0.9
I132V	2.0 ± 0.1	0.8
I276V	1.7 ± 0.2	0.7
V291I	2.3 ± 0.1	0.9
V361A	3.6 ± 0.1	1.4
A409V	1.9 ± 0.1	0.7
A411T	3.1 ± 0.1	1.2
A413V	2.0 ± 0.1	0.8
R417Q	0.47 ± 0.1	0.2
D429N	2.7 ± 0.1	1.1

^aValues represent means \pm SD from at least three experiments.

^bUnits are M DNA cleaved/(min \times M enzyme).

^cRelative activity is the ratio of the DH1-1A or singly substituted PI-*SceI* to *S.cerevisiae* values.

assayed protein splicing activity. Protein splicing of this fusion protein, termed MYB (maltose-binding protein-yeast intein-chitin-binding domain), normally yields the excised VMA intein and a second protein comprised of the fused maltose-binding protein and chitin-binding domain. Two variants were constructed to eliminate the contact (R44A/D429 and R44/D429A) and one was made to invert the interacting residues (R44D/D429R). The residue pairing observed in DH1-1A was constructed (R49S/D429N) as well as intermediates that contain only one or other of the substitutions (R44S/D429 and R44/D429N). Western blot analysis of cultures grown containing each of these fusion derivatives using anti-intein and anti-maltose-binding protein antibodies revealed that all of the proteins spliced like the wild-type construct (data not shown). Therefore, it is unlikely that the Arg44-Asp429 salt bridge plays a key role in maintaining the conformation and function of the protein splicing active site.

DISCUSSION

A 'life cycle' of mobile inteins and introns and their resident HEG has been proposed that describes recurrent cycles of invasion, maintenance and eventual deletion (33). In the first stage, the mobile intein/intron invades, presumably by horizontal transmission, the locus of a new genome that is not already a host to the element. This gene transfer occurs between organellar and nuclear genomes within one organism or between species genomes through the aid of a vectoring agent (plasmid, virus, etc.) or by other means. Once situated in its new genome, it is transmitted by vertical transmission to successive generations and is stably maintained in the population by homing, which ensures that uninfected individuals acquire the element at the same locus. The absence of a positive selection

for the element, however, eventually leads to genetic drift and to its degeneration. Loss of the HEG activity through deletion or point mutation would be expected to have minimal effects on the host and is likely to occur prior to loss of the RNA or protein splicing activities of the intron/intein, which are necessary for production of the host protein. If the HEG activity is completely lost, for example by mutation of catalytic residues, homing will cease and the HEG may eventually be deleted. Precise deletion of the intron or intein protein splicing component will ensue, resulting in a genome that is devoid of the entire mobile element. However, HEGs tolerate mutation of their DNA-binding surface residues and still maintain the ability to cleave the same or similar recognition sequences (34), even if at reduced efficiency. Consequently, it is possible that prior to deletion these elements are transferred by horizontal transmission to a new host genome containing the recognition site for the altered enzyme, and the cycle begins anew.

The evidence for lateral transmission of these elements has come primarily from phylogenetic comparisons. In studies of the ω homing intron from *S.cerevisiae* (33) and of the mobile group I intron from the mitochondrial *cox1* gene in angiosperm plants (35) the observed incongruence between the phylogenies of the host organism and of the mobile intron/HEG element provides evidence of numerous episodes of lateral transmission. Mobility of group I intron/intein homing elements may involve their insertion into double-strand breaks created by the homing endonuclease. This has not been experimentally demonstrated, but it has been shown that DNAs derived from yeast Ty retrotransposons or from the mitochondrial genome fill double-strand breaks in the nuclear genome when no DNA homologs are available (36,37). Moreover, invasion of a homing group II intron into ectopic chromosomal sites by a retrotransposition mechanism has been reported (38). Degeneration of the LAGLIDADG homing endonucleases by deletion or insertion mutations is readily apparent in comparisons of the ω element from different yeast strains and of LAGLIDADG inteins from different organisms (33,39). The extent of deletion varies, but in some cases the entire HEG region of an intein or intron is absent. In this regard, a comparison of the X-ray structures of the *M.xenopi* GyrA mini-intein, which lacks HEG motifs, and the *S.cerevisiae* VMA intein reveals that the entire homing endonuclease domain is absent from the mini-intein (9).

In this report, we provide a detailed molecular analysis of an intein and of a target site derived from a tetraploid yeast strain used for wine making. The co-existence of both donor and recipient alleles in the same strain is surprising and may be directly related to its origin. Most wine making and brewer's yeasts are polyploids that arose as interspecific diploid hybrids between *S.cerevisiae* and a closely related species (i.e. *Saccharomyces paradoxus* or *Saccharomyces bayanus*) (40,41). The origin of DH1-1A is unknown, but in one scenario, the strain arose as a naturally occurring interspecific hybrid of sibling species of *Saccharomyces*, one containing an intein⁺ and one containing an intein⁻ allele. Since the predicted target sequence from the intein⁺ allele exactly matches the PI-SceI site in *S.cerevisiae*, this allele may have originally arisen by an intein homing event initiated by a PI-SceI-like enzyme. In contrast, the DH1-1A intein⁻ recognition sequence contains seven mutations that only permit it to be cleaved by PI-SceI under relaxed conditions, and this reduced activity may explain

why this allele has not undergone gene conversion. Whether these mutations occurred before or after association of the intein⁺ and intein⁻ alleles is uncertain, but we favor the former possibility since we expect that the intein⁻ allele would be converted to an intein⁺ allele if it were completely functional in the presence of an active homing endonuclease. Alternatively, the 11 mutations in the DH1-1A enzyme that substantially inactivate the homing endonuclease may already have been present when the two alleles associated, thereby preventing gene conversion of the intein⁻ allele. The inability to transit through meiosis, which is when homing occurs, may also explain the existence of the intein⁻ allele (5).

It is likely that substantial degeneration of the intein⁺ and intein⁻ alleles has occurred since the DNA cleavage activities of the DH1-1A target site and the homing endonuclease are significantly less than their *S.cerevisiae* counterparts. The cleavage defects of the DH1-1A intein⁻ and *C.tropicalis* target sequences most likely result from the cumulative effect of their mutations because none occurs at any of the critical positions. However, in spite of the fact that nearly a quarter of the base pairs within the recognition sequence are changed, the substrates are still cleaved *in vitro*, albeit poorly, by PI-SceI, which supports the general finding that homing endonucleases tolerate significant target site degeneracy (20,42–49). Also noteworthy is the observation that none of the mutations in either the DH1-1A or the *C.tropicalis* substrates interferes with the ATPase subunit coding frame, suggesting that the sequence of the protein in this region, which includes an essential catalytic residue (Glu286), is strongly selected. In the case of inteins, the host protein sequence imposes limits on the sequence divergence of the homing endonuclease recognition site.

The small number of differences between the DH1-1A homing endonuclease and *S.cerevisiae* PI-SceI (97.6% identity) contrasts sharply with the low sequence identity between the DH1-1A and *C.tropicalis* intein sequences (35.9% identity) (21), suggesting that DH1-1A and *S.cerevisiae* are more closely related than DH1-1A and *C.tropicalis*. Protein splicing activity is maintained in both cases even when the endonuclease activity is lost. A single substitution at Arg417 accounts for most of the defect in DNA cleavage and binding of the DH1-1A endonuclease. This result is consistent with the finding that substitution of Arg417 with alanine reduces DNA cleavage and DNA binding ~5-fold (13). The Arg417 mutation reduces the PI-SceI DNA binding and cleavage activities less than that of other mutations that have been characterized in the protein (11,13,32). Inspection of the model of the PI-SceI–DNA complex positions this residue near the DNA substrate (50), very close to where the neighboring residue Tyr420 crosslinks to the phosphate backbone at position +12/+13 (51). Whether Arg417 directly or indirectly contacts the DNA or whether it affects DNA binding through other residues cannot be determined from this study.

We observed that two residues that normally form a surface salt bridge, Arg44 and Asp429, are replaced by Ser44 and Asn429 in the DH1-1A protein, which are capable of forming a hydrogen bond. One possibility is that sequential complementary substitutions occurred at these positions in the protein in order to maintain an essential function. We show here that in spite of the fact that the Ser44–Asn429 interaction is proximal to the protein splicing active site, mutation of these residues affects neither the endonuclease nor the protein splicing activities.

The possibility that the salt bridge in PI-*SceI* and the putative hydrogen bond in the DH1-1A endonuclease influences protein stability cannot be ruled out, but has not been tested here.

The eventual fate of the DH1-1A intein could take several forms. In one scenario, it accumulates numerous inactivating point mutations, which is followed by deletion of the HEG and/or protein splicing regions of the gene. The *Candida* HEG may have already started along this pathway since the two conserved acidic residues at the catalytic center are replaced by isoleucine and alanine (21), which presumably inactivate the enzyme. However, the endonuclease may also begin the invasion cycle anew if, prior to its complete inactivation, it is transferred to a new genomic environment where its accumulated mutations enable it to cleave a related site. In a recent study of 45 single LAGLIDADG proteins related to I-*CreI* from different green algae strains it is apparent that these proteins recognize similar DNA recognition sequences using different subsets of residues within the protein (34). Therefore, identical protein-DNA contacts may not be maintained in a subfamily of LAGLIDADG enzymes. Just as in the case of the DH1-1A endonuclease, these algal enzymes sometimes cleave their DNA substrates less efficiently (34), which may indicate that some degeneration of these HEGs has already occurred. Thus, degeneration is tolerated by LAGLIDADG enzymes, but eventually it will lead to inactivation of the endonuclease and to HEG elimination.

ACKNOWLEDGEMENTS

We thank Dr I. Herskowitz for his gift of yeast strains and Dr R. Sinden for the use of his Kodak EDAS 290 imager. We thank Dr Karen Posey for comments on the manuscript and for insightful discussions. We acknowledge Dr J. Thorner in whose laboratory this work was initiated. This work was supported by a grant from the Welch Foundation.

REFERENCES

- Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA and genome evolution. *Evolution*, **55**, 1–24.
- Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
- Jurica, M.S. and Stoddard, B.L. (1999) Homing endonucleases: structure, function and evolution. *Cell. Mol. Life Sci.*, **55**, 1304–1326.
- Gimble, F.S. (2000) Invasion of a multitude of genetic niches by homing endonuclease genes. *FEMS Microbiol. Lett.*, **185**, 99–107.
- Gimble, F.S. and Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, **357**, 301–306.
- Kuhlmann, U.C., Moore, G.R., James, R., Kleanthous, C. and Hemmings, A.M. (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.*, **463**, 1–2.
- Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W.R., Chatterjee, A. and Mian, I.S. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.*, **25**, 4626–4638.
- Duan, X., Gimble, F.S. and Quiocho, F.A. (1997) Crystal structure of PI-*SceI*, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555–564.
- Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R. and Sacchettini, J.C. (1998) Crystal structure of *gyrA* intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Struct. Biol.*, **5**, 31–36.
- Hall, T.M.T., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. and Leahy, D.J. (1997) Crystal structure of a hedgehog autoprocessing domain: homology between hedgehog and self-splicing proteins. *Cell*, **91**, 85–97.
- He, Z., Crist, M., Yen, H.-C., Duan, X., Quiocho, F.A. and Gimble, F.S. (1998) Amino acid residues in both the protein splicing and endonuclease domains of the PI-*SceI* intein mediate DNA binding. *J. Biol. Chem.*, **273**, 4607–4615.
- Grindl, W., Wende, W., Pingoud, V. and Pingoud, A. (1998) The protein splicing domain of the homing endonuclease PI-*SceI* is responsible for specific DNA binding. *Nucleic Acids Res.*, **26**, 1857–1862.
- Wende, W., Schottler, S., Grindl, W., Christ, F., Steuer, S., Noel, A.J., Pingoud, V. and Pingoud, A. (2000) Analysis of binding and cleavage of DNA by the gene conversion PI-*SceI* endonuclease using site-directed mutagenesis. *Mol. Biol. (Mosk.)*, **34**, 1054–1064.
- Gimble, F.S. and Stephens, B.W. (1995) Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of PI-*SceI* endonuclease. *J. Biol. Chem.*, **270**, 5849–5856.
- Schöttler, S., Wende, W., Pingoud, V. and Pingoud, A. (2000) Identification of Asp218 and Asp326 as the principal Mg²⁺ binding ligands of the homing endonuclease PI-*SceI*. *Biochemistry*, **39**, 15895–15900.
- Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. and Morikawa, K. (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-*PfuI*. *J. Mol. Biol.*, **300**, 889–901.
- Heath, P.J., Stephens, K.M., Monnat, R.J. Jr and Stoddard, B.L. (1997) The structure of I-*CreI*, a group I intron-encoded homing endonuclease. *Nature Struct. Biol.*, **4**, 468–476.
- Silva, G.H., Dalgaard, J.Z., Belfort, M. and Van Roey, P. (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-*DmoI*. *J. Mol. Biol.*, **286**, 1123–1136.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Gimble, F.S. and Wang, J. (1996) Substrate recognition and induced DNA distortion by the PI-*SceI* endonuclease, an enzyme generated by protein splicing. *J. Mol. Biol.*, **263**, 163–180.
- Gu, H.H., Xu, J., Gallagher, M. and Dean, G.E. (1993) Peptide splicing in the vacuolar ATPase subunit A from *Candida tropicalis*. *J. Biol. Chem.*, **268**, 7372–7381.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene (Amst.)*, **77**, 51–59.
- Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A. et al. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene (Amst.)*, **192**, 271–281.
- Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F.B. and Xu, M.-Q. (1996) Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. *J. Biol. Chem.*, **271**, 22159–22168.
- Bowman, E.J., Tenney, K.T. and Bowman, B.J. (1988) Isolation of genes encoding the *Neurospora* vacuolar ATPase. *J. Biol. Chem.*, **263**, 13994–14001.
- Ghislain, M. and Bowman, E.J. (1992) Sequence of the genes encoding subunits A and B of the vacuolar H⁺-ATPase of *Schizosaccharomyces pombe*. *Yeast*, **8**, 791–799.
- Forster, C., Santos, M.A., Ruffert, S., Kramer, R. and Revuelta, J.L. (1999) Physiological consequence of disruption of the VMA1 gene in the riboflavin overproducer *Ashbya gossypii*. *J. Biol. Chem.*, **274**, 9442–9448.
- Liu, J. and Kane, P.M. (1996) Mutational analysis of the catalytic subunit of the yeast vacuolar proton-translocating ATPase. *Biochemistry*, **35**, 10938–10948.
- Liu, Q., Leng, X.H., Newman, P.R., Vasilyeva, E., Kane, P.M. and Forgac, M. (1997) Site-directed mutagenesis of the yeast V-ATPase A subunit. *J. Biol. Chem.*, **272**, 11750–11756.
- MacLeod, K.J., Vasilyeva, E., Baleja, J.D. and Forgac, M. (1998) Mutational analysis of the nucleotide binding sites of the yeast vacuolar proton-translocating ATPase. *J. Biol. Chem.*, **273**, 150–156.
- Wende, W., Grindl, W., Christ, F., Pingoud, A. and Pingoud, V. (1996) Binding, bending and cleavage of DNA substrates by the homing endonuclease PI-*SceI*. *Nucleic Acids Res.*, **24**, 4123–4132.

32. Hu, D., Crist, M., Duan, X. and Gimble, F.S. (1999) Mapping of a DNA binding region of the *PI-SceI* homing endonuclease by affinity cleavage and alanine-scanning mutagenesis. *Biochemistry*, **38**, 12621–12628.
33. Goddard, M.R. and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl Acad. Sci. USA*, **96**, 13880–13885.
34. Lucas, P., Otis, C., Mercier, J.P., Turmel, M. and Lemieux, C. (2001) Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. *Nucleic Acids Res.*, **29**, 960–969.
35. 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.
36. Moore, J.K. and Haber, J.E. (1996) Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature*, **383**, 644–646.
37. Yu, X. and Gabriel, A. (1999) Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell*, **4**, 873–881.
38. Cousineau, B., Lawrence, S., Smith, D. and Belfort, M. (2000) Retrotransposition of a bacterial group II intron. *Nature*, **404**, 1018–1021.
39. Perler, F.B., Olsen, G.J. and Adam, E. (1997) Compilation and analysis of intein sequences. *Nucleic Acids Res.*, **25**, 1087–1094.
40. Hammond, J.R.M. (1993) Brewer's yeasts. In Rose, A. and Harrison, S. (eds), *The Yeasts*, 2nd Edn. Academic Press, New York, NY, Vol. 5, pp. 8–11.
41. Kielland-Brandt, M.C., Nilsson-Tillgren, T., Gjermansen, C., Holmberg, S. and Pedersen, M.B. (1995) Genetics of brewing yeasts. In Wheals, A., Rose, A. and Harrison, S. (eds), *The Yeasts*, 2nd Edn. Academic Press, New York, NY, Vol. 6.
42. Colleaux, L., D'Auriol, L., Galibert, F. and Dujon, B. (1988) Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl Acad. Sci. USA*, **85**, 6022–6026.
43. Marshall, P. and Lemieux, C. (1992) The *I-CeuI* endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene. *Nucleic Acids Res.*, **20**, 6401–6407.
44. Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S. and Butow, R.A. (1992) Complex recognition site for the group I intron-encoded endonuclease *I-SceII*. *Mol. Cell. Biol.*, **12**, 716–723.
45. Bryk, M., Quirk, S.M., Mueller, J.E., Loizos, N., Lawrence, C. and Belfort, M. (1993) The *td* intron endonuclease *I-TevI* makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J.*, **12**, 2141–2149.
46. Durrenberger, F. and Rochaix, J.D. (1993) Characterization of the cleavage site and the recognition sequence of the *I-CreI* DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.*, **236**, 409–414.
47. Aagaard, C., Awayez, M.J. and Garrett, R.A. (1997) Profile of the DNA recognition site of the archaeal homing endonuclease *I-DmoI*. *Nucleic Acids Res.*, **25**, 1523–1530.
48. Argast, G.M., Stephens, K.M., Emond, M.J. and Monnat, R.J.Jr (1998) *I-PpoI* and *I-CreI* homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.*, **280**, 345–353.
49. Wittmayer, P.K., McKenzie, J.L. and Raines, R.T. (1998) Degenerate DNA recognition by *I-PpoI* endonuclease. *Gene (Amst.)*, **206**, 11–21.
50. Hu, D., Crist, M., Duan, X., Quiocho, F.A. and Gimble, F.S. (2000) Probing the structure of the *PI-SceI*-DNA complex by affinity cleavage and affinity photocross-linking. *J. Biol. Chem.*, **275**, 2705–2712.
51. Christ, F., Steuer, S., Thole, H., Wende, W., Pingoud, A. and Pingoud, V. (2000) A model for the *PI-SceI*-DNA complex based on multiple base and phosphate backbone-specific photocross-links. *J. Mol. Biol.*, **300**, 867–875.
52. Christopher, J.A. (1998) *SPOCK Software*, version 1.0b135. Texas A&M University, College Station, TX.