Homology Modeling and Mutational Analysis of Ho Endonuclease of Yeast

Anya Bakhrat,* Melissa S. Jurica,[†,1] Barry L. Stoddard[†] and Dina Raveh*[,2]

**Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, 84105 Israel and [†]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109*

Manuscript received August 7, 2003
Accepted for publication October 31, 2003

ABSTRACT

Ho endonuclease is a LAGLIDADG homing endonuclease that initiates mating-type interconversion in yeast. Ho is encoded by a free-standing gene but shows 50% primary sequence similarity to the intein (protein-intron encoded) PI-*Sce*I. Ho is unique among LAGLIDADG endonucleases in having a 120-residue C-terminal putative zinc finger domain. The crystal structure of PI-*Sce*I revealed a bipartite enzyme with a protein-splicing domain (Hint) and intervening endonuclease domain. We made a homology model for Ho on the basis of the PI-*Sce*I structure and performed mutational analysis of putative critical residues, using a mating-type switch as a bioassay for activity and GFP-fusion proteins to detect nuclear localization. We found that residues of the N-terminal sequence of the Hint domain are important for Ho activity, in particular the DNA recognition region. C-terminal residues of the Hint domain are dispensable for Ho activity; however, the C-terminal putative zinc finger domain is essential. Mutational analysis indicated that residues in Ho that are conserved relative to catalytic, active-site residues in PI-*Sce*I and other related homing endonucleases are essential for Ho activity. Our results indicate that in addition to the conserved catalytic residues, Hint domain residues and the zinc finger domain have evolved a critical role in Ho activity.

HO endonuclease initiates a mating-type switch in the yeast *Saccharomyces cerevisiae* by making a double-strand break (DSB) in a 24-bp cognate sequence in the mating-type gene *MAT* (Kostriken and Heffron 1984). Repair of the DSB is by gene conversion using one of the silent *HM* cassettes as a template and results in substitution of the resident *MAT* allele with a sequence of the opposite mating type (Strathern *et al.* 1982; Simon *et al.* 2002). Ho has homology to LAGLI DADG homing endonucleases (Neff 1993), rare-cutting enzymes that cleave long (>14–40 bp) asymmetrical target sequences in the minor groove, leaving 4-nucleotide (nt) 3′ cohesive ends. Ho (F-*Sce*II) is encoded by a freestanding nuclear gene; however, other homing endonucleases are encoded by open reading frames (ORF) embedded in genetically mobile introns or within self-splicing inteins [protein introns (PI); Belfort and Roberts 1997; Jurica and Stoddard 1999; Kowalski and Derbyshire 2002]. Repair of the DSB made by the endonuclease promotes mobility of its host intron/intein into its cognate site. The survival of Ho has been attributed to its ability to promote a mating-type switch, thus allowing progeny of a single haploid spore to mate and sporulate (Gimble 2000). Eight conserved sequence motifs are found in intein-encoded LAGLIDADG endonucleases, and the primary sequence of Ho displays all the intein motifs, including those involved in protein splicing (Pietrokovski 1994; Perler 2000).

Structures have been determined using X-ray crystallography for a number of LAGLIDADG endonucleases (Chevalier and Stoddard 2001). The *Chlamydomonas reinhardtii* 23S rRNA intron-encoded I-*Cre*I has a single copy of the LAGLIDADG motif and acts as a homodimer, cleaving an almost palindromic homing site (Heath *et al.* 1997; Jurica *et al.* 1998; Chevalier *et al.* 2001). The αββαββα fold of the I-*Cre*I subunit is found in two copies in the monomeric endonucleases, yeast PI-*Sce*I (Duan *et al.* 1997; Hu *et al.* 2000), I-*Dmo*I (Silva *et al.* 1999), and PI-*Pfu*I (Ichiyanagi *et al.* 2000). These monomeric forms are thought to have arisen by gene duplication (Lykke-Andersen *et al.* 1996; Silva *et al.* 1999). In all the structures the two LAGLIDADG α-helices pack against each other with a pair of catalytic aspartic acid residues at the C-terminal ends of each helix. The LAGLIDADG α-helix is followed by two anti-parallel β-strands with a loop of varying length between them. The length of the β-strands and size of the intervening loop are correlated with cognate site length, whereas individual residues in this region dictate site specificity (Jurica and Stoddard 1999; Chevalier and Stoddard 2001). All LAGLIDADG enzyme structures solved to date show the central two-helix bundle followed by two β-strands and the αββαββα fold represents the domain topology for the family (Jurica and Stoddard 1999; Chevalier and Stoddard 2001).

[1]*Present address:* Howard Hughes Medical Institute, Department of Biochemistry, Brandeis University, Waltham, MA 02454.

[2]*Corresponding author:* Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, 84105 Israel.
E-mail: raveh@bgumail.bgu.ac.il

The primary sequence of Ho has ~50% similarity to PI-*Sce*I (Hirata *et al.* 1990). The crystal structure of PI-*Sce*I shows two distinct domains: Domain I corresponds to the protein-splicing activity and is composed of residues from both the N and the C terminus of the primary sequence; domain II is an intervening endonuclease domain centered around the two LAGLIDADG α-helices with a structure resembling that of the I-*Cre*I homodimer (Duan *et al.* 1997; Hu *et al.* 2000). Within domain I protein-splicing activity is contained in a horseshoe-like structural motif known as the Hedgehog and intein (Hint) domain (Hall *et al.* 1997; Hu *et al.* 2000). In PI-*Sce*I an additional DNA recognition region (DRR) is present in domain I where residues important for DNA binding have been identified (He *et al.* 1998; Moure *et al.* 2002). PI-*Sce*I domain I expressed by itself in *Escherichia coli* is able to bind a PI-*Sce*I cognate site oligomer whereas isolated endonuclease domain II does not bind (Grindl *et al.* 1998). In addition to PI-*Sce*I a DRR can be identified in the primary sequence of Ho and of PI-*Ctr*I of *Candida tropicalis*. PI-*Pfu*I has a unique stirrup fold inserted between the Hint and endonuclease domains that may be analogous (Ichiyanagi *et al.* 2000).

Ho is the only member of this large family of >200 homing endonucleases that has a C-terminal zinc-binding domain with a putative role in cognate site recognition (Russell *et al.* 1986). We therefore undertook a functional analysis of Ho to identify residues important for its activity, using a mating-type switch as a bioassay for the ability of the enzyme to make a site-specific DSB at the *MAT* locus. We created a homology model of Ho on the basis of the coordinates of PI-*Sce*I, using the program MODELLER (Molecular Structure) and tested the importance of residues of the protein-splicing domain for Ho activity. We superposed the model onto four solved LAGLIDADG endonuclease structures, I-*Cre*I, I-*Dmo*I, PI-*Sce*I, and PI-*Pfu*I, to identify putative active site residues that were then tested by site-directed mutagenesis. Furthermore we analyzed the unique zinc finger domain of Ho and determined the borders of this region.

MATERIALS AND METHODS

Yeast plasmids and strains: *HO* was cloned into pYES2 (Invitrogen, San Diego) as a *Hin*dIII fragment starting 171 bp upstream of the first ATG. Single-residue substitutions were made with the Quikchange kit (Stratagene, La Jolla, CA). Each mutation introduced a novel diagnostic restriction site and was confirmed by direct sequencing. Primers are listed in Table 1. The plasmids were transformed into a Δ*ho* strain of yeast, JKM120: *HML*α, *MAT***a** *HMR***a**, Δ*ho, ade-1, leu 2-3,112, lys5, trp1::hisG, ura3-52*. The *ho* deletion removes the promoter and 705 bp of *ho* sequence (J. K. Moore and J. E. Haber, unpublished results). Mutant *HO* genes were cloned as green fluorescent protein (GFP) fusions in pHY315-GFP (Kaplun *et al.* 2003), using primers HoSpeF and HoSacR. The 54-bp C-terminal truncation employed primer pHYTruncR. The fusion proteins were expressed in Δ*msn5* mutant cells: W303 background, Δ*leu2, ura3-52 his3-200 msn5::HIS3*, obtained from J. Hood.

Ho activity assays: *Southern analysis:* Cell cultures (50 ml) were grown overnight in minimal medium with 2% raffinose. *HO* expression was induced by addition of 2% galactose for 3 hr. Five milliliters were taken for a mating-type test, and the rest was used for DNA extraction. The DNA was digested with *Hin*dIII and run on a 0.8% agarose gel for blotting. A radioactive probe was prepared by random primer extension using a 4.3-kb *MAT* fragment as a template (Maniatis *et al.* 1982). Ho cleavage of *MAT* gives two bands of ~3.5 and 1.5 kb.

Mating-type test: pYES-*HO* and the mutated versions were transformed into JKM120 that is *MAT***a**. Induced and noninduced cells were mated with a *MAT*α tester to determine their mating efficiency. The activity of the *HO* genes was assayed by measuring the ability of the *MAT***a** transformants to mate with a *MAT***a** tester strain and give rise to diploids on minimal plates. Mating tester strains were DC14 (*MAT***a**, *his1*) and DC17 (*MAT*α, *his1*). Standard techniques and media composition were from Sherman *et al.* (1986).

RESULTS

Homology model of Ho based on the structure of PI-*Sce*I: The sequences of Ho and PI-*Sce*I are ~50% similar with few gaps in the alignment, which was helpful in our model building. Ho and PI-*Sce*I have 29% identity across the intein domain and 35% identity across the endonuclease domain, the highest similarity being in the LAGLIDADG motifs (Figure 1).

We constructed a homology model of Ho on the basis of the coordinates of the PI-*Sce*I apoprotein as a template and a primary sequence alignment of Ho with PI-*Sce*I made using GAP of the Wisconsin Genetics Computer Group program (Wisconsin Package Version 10.0). Side chains in the PI-*Sce*I structure were mutated to the Ho sequence on the basis of this alignment using the program QUANTA (Molecular Simulations). Insertions and deletions in the alignment were also incorporated into the mutated structure. The mutated structure together with the parent PI-*Sce*I structure and the sequence alignment were used as input for the program MODELLER (Molecular Structure). This program performs an energy minimization of the mutated structure while employing constraints to the atomic coordinates of aligned residues in the parent structure to generate a refined homology model.

The Ho homology model includes residues 1–465 with the exception of residues 98–108 that correspond to a disordered loop in the PI-*Sce*I structure (residues 93–102). Initially these residues were present in the homology model, but failed to converge during the minimization process due to the lack of structural information in the PI-*Sce*I parent structure and were therefore removed from the final model. Two other disordered loops in the PI-*Sce*I structure (residues 271–279 and 369–374) were included in the Ho model with poorer but acceptable geometry.

Despite being encoded by a free-standing gene, the primary sequence of Ho contains the Hint sequence

FIGURE 1.—Intein motifs of PI-*Sce*I and Ho. Sequence alignment of PI-*Sce*I and Ho showing Hint sequence motifs A, N2, B, F, and G; endonuclease motifs C, D, E, and H; and the unique C-terminal zinc binding domain of Ho. The DNA recognition region (DRR) of PI-*Sce*I is boxed as are the C-X2-C motifs of the three putative zinc fingers.

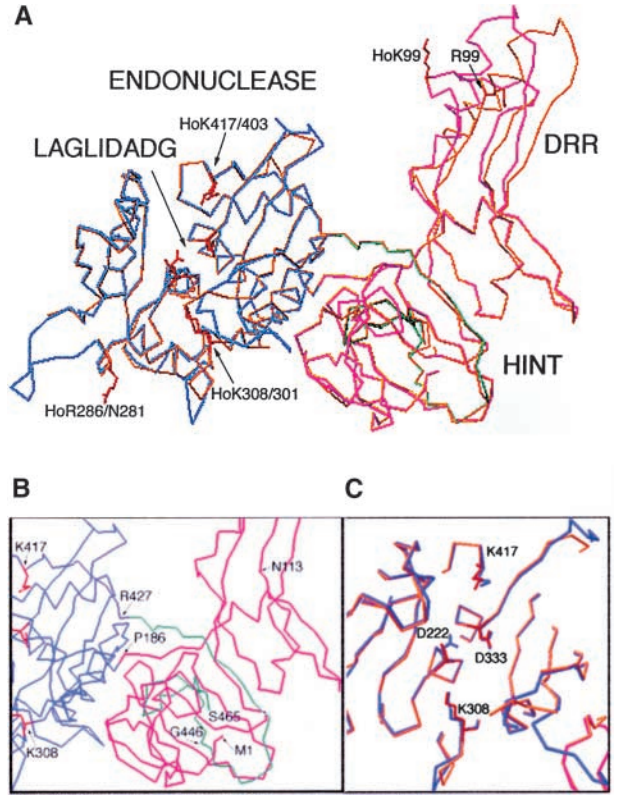

FIGURE 2.—(A) Superposition of Ho homology model onto PI-*Sce*I structure; the Hint and DRR that comprise the protein splicing domain are indicated, as is the endonuclease domain. The backbone of PI-*Sce*I is in orange, the N-terminal domain I residues of Ho are in pink, the endonuclease residues are in blue, and the C-terminal domain I residues are in green. Residues of Ho that were mutated are indicated with red side chains and are labeled; numbers of parallel residues of PI-*Sce*I appear after the slash. The zinc finger domain of Ho is not shown. (B) Hint domain and part of DRR extension of Ho with residues M1 and S465 that delineate the borders of the model indicated. P186 and R427 are the borders of the intervening endonuclease domain. Also marked are residues that served for truncations and deletions described in the text. (C) Slab showing the side chains of the active site residues of PI-*Sce*I (red) and of Ho (blue) that were mutated. Numbering of residues is as for Ho.

motifs, with inactivating mutations in critical functional residues (PERLER *et al.* 1997), in addition to the endonuclease domain II (Figure 2A). Domain I of Ho aligns well with the three-dimensional structure of PI-*Sce*I and is composed of 185 N-terminal residues encompassing intein sequence motifs A, N2, and B. The C-terminal part extends from Glu428 to Ser465 and includes intein sequence motifs F and G. The endonuclease domain, characterized by two repeats of the αββαββα fold of the I-*Cre*I subunit, begins at approximately Pro186 and ends at Arg427; it includes intein sequence motifs C, D, E, and H (Figures 1 and 2). The sequence downstream of domain II is composed of intein motifs F and G that form the C-terminal portion of the Hint fold. Downstream of this sequence is the unique 120-residue zinc finger domain. This region of Ho has no homology to PI-*Sce*I and was not included in the modeling.

The root mean square deviation between the PI-*Sce*I structure and Ho model is 0.365 Å^2. The largest differences correlate to insertions in the Ho sequence at loops in the endonuclease domain (Figure 2A). There is a two-residue insertion preceding block D of the conserved intein motifs, which contains a proposed catalytic residue (Lys308). The conformation of these inserted residues appears perturbed, and it is likely that in the native structure a change in the loop structure preceding this insertion would be observed.

Dissection of the role of the domain I residues in Ho activity: We deleted the N-terminal 112 residues of Ho and found that this led to loss of activity. This contradicts an earlier finding of activity for the truncated gene (NAHON and RAVEH 1998). In PI-*Sce*I residues Arg90 and Arg94 of the DRR have a major role in cognate site binding (HE *et al.* 1998). The model places Ho Lys99 in the vicinity of these residues of PI-*Sce*I (Figure 2A) and we therefore mutated Lys99 to alanine. We found that this abrogated the ability of Ho to initiate a mating-type switch. The truncation and the mutation indicate

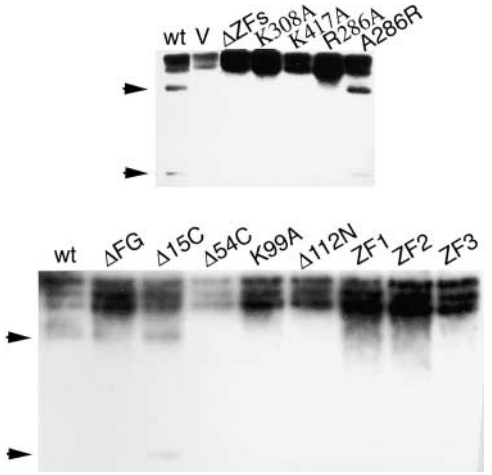

FIGURE 3.—Southern analysis to test for activity of Ho mutants. (Top) wt, cleavage of *MAT* after induction of pYES-*HO* into two bands of 3.5 and 1.5 kb (arrows); V, pYES control; ΔZFs, C-terminal truncation of Ho—remaining protein consisting of residues 1–446 has no endonucleolytic activity; K308A substitution; K417A substitution; R286A substitution; and Ala286 restored to arg, the restored allele shows cleavage of *MAT*. (Bottom) wt; ΔFG is an internal deletion of the C-terminal F and G intein motifs, residues 446–465, and extends to residue 586; Δ15 is a C-terminal truncation of 15 residues and extends to residue 571; Δ54, of 54 C-terminal residues to residue 532; Δ112N, of 112 N-terminal residues; and ZF1, ZF2, and ZF3 are substitutions of the upstream two cysteine residues of each putative zinc finger motif with alanine residues.

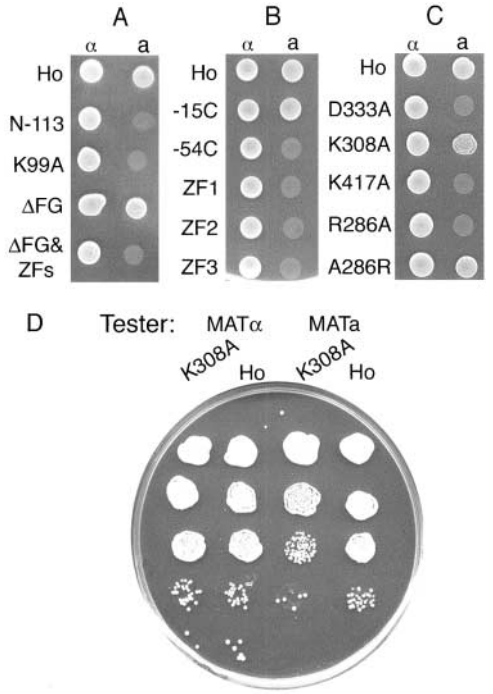

FIGURE 4.—Mating-type tests of Ho and mutant versions. *MAT***a** cells transformed with a specific *HO* plasmid were mated with the *MAT*α tester control and then tested for their ability to mate with the *MAT***a** tester strain. (A) Ho; N-113, N-terminal truncation of 113 residues; K99A in the DRR; ΔFG, internal deletion of intein motifs F and G (residues 446–465); ΔFG and ZFs, C-terminal truncation of intein motifs F and G and the zinc finger domain (remaining endonuclease extends from 1 to 446). Ho and ΔFG show mating with the *MAT***a** tester. (B) Ho; −15C, C-terminal deletion of 15 residues; −54C, C-terminal deletion of 54 residues; ZF1, mutations of C466A and C469A; ZF2, C508A and C511A mutations; ZF3, C558A and C561A substitutions. Ho and the 15-residue deletion show mating with the *MAT***a** tester. (C) Ho, D333A, K308A, K417A, R286A, and A286R restored. Ho and the restored R286 show mating, and K308A shows residual mating ability. (Plate) Serial 10-fold dilutions of *MAT***a** cells transformed with column 1, Ho-K308A, and column 2, Ho, were mated with the control *MAT*α tester. Both cells show the same mating efficiency. In columns 3 and 4 Ho-K308A and Ho, respectively, were mated with the *MAT***a** tester. The K308A mutation reduces the mating-type switching efficiency of Ho ~10-fold.

that in Ho as in PI-*Sce*I, domain I residues are important for site-specific DNA cleavage (Figures 3 and 4).

In the Ho model the C-terminal part of domain I starts at Arg427, 56 residues downstream of domain II intein motif H, and ends at Ser465 before the first Cys of the putative zinc finger domain (Figure 2B). It includes intein sequence motifs F and G. We truncated both these sequence motifs and the zinc finger domain to make a protein that extends from residue 1 to Gly446 (Figure 2B). This C-terminal domain-truncated Ho displays no activity. However, when we restored the zinc finger domain by making an internal deletion of motifs F and G (residues 446–465) we found that this construct could induce a mating-type switch (Figures 3–5).

Dissection of the zinc finger domain of Ho: The zinc finger structure is stabilized by a zinc ion bound to four cysteine moieties and three linearly ordered zinc fingers can be predicted from the primary sequence of the zinc-binding domain of Ho (RUSSELL *et al.* 1986). The first two putative zinc fingers of Ho correspond to the C2C2-C2C2 zinc finger type and have a sequence Cys-X2-Cys-X16-Cys-X2-Cys-X18-Cys-X2-Cys-X10-Cys-X2-Cys (KLUG and SCHWABE 1995). They are followed by 32 residues that precede a putative third zinc finger of sequence C-X2-C-X12-H-X2-C that terminates 9 residues before the C terminus of Ho. We constructed a 15-residue C-terminal truncation of Ho that eliminates half the putative third zinc finger. This truncated form of Ho retains endonucleolytic activity. The zinc finger domain is, however, essential for Ho activity as a 54-residue truncation of Ho that eliminates the entire third zinc finger and 26 upstream residues caused Ho to lose its activity (Figures 3–5). We therefore substituted two alanine residues for the upstream cysteines of each putative zinc finger. Mutation of the cysteine residues of each of the three putative zinc fingers resulted in loss of Ho activity (Figures 4 and 5).

We have shown previously that Ho is a very unstable protein with a half-life of ~8 min (KAPLUN *et al.* 2000). To ensure that mutant versions of Ho that had lost the capability to induce a mating-type switch were stable and

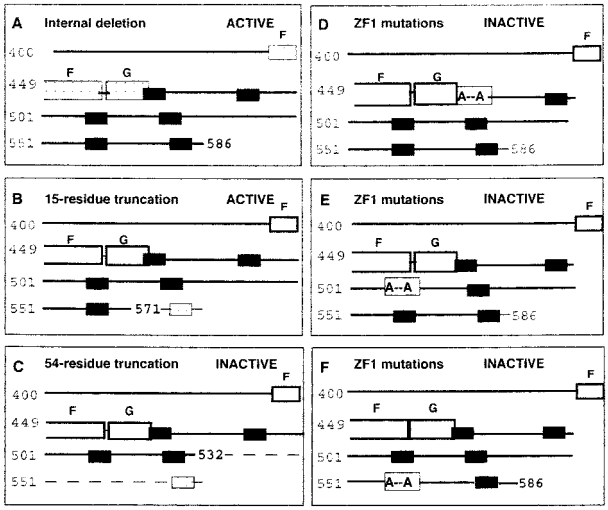

FIGURE 5.—Diagram showing the results of mutations and deletions in the C terminus of Ho. (A) Internal deletion of intein sequence motifs F and G does not affect Ho activity. (B) A 15-residue C-terminal truncation does not affect activity. (C) A 54-residue C-terminal truncation abolishes activity. (D) First zinc finger C466A and C469A mutations inactivate Ho. (E) Second zinc finger C508A and C511A mutations inactivate. (F) Third zinc finger C558A and C561A substitutions inactivate.

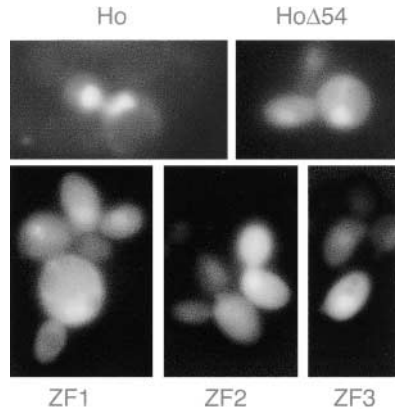

FIGURE 6.—Localization of C-terminal zinc finger mutations and deletions of Ho using GFP-fusion proteins expressed in cells deleted for the Msn5 nuclear exportin. HoΔ54 is the C-terminal deletion to residue 532; the zinc fingers are numbered from the N terminus as in Figure 5.

were imported into the nucleus, we fused the inactive 54-residue C-terminal truncation and the three zinc finger mutants to GFP to determine their nuclear accumulation. GFP-fusion proteins were expressed in cells deleted for the Msn5 nuclear exportin as in these mutants Ho remains within the nucleus (KAPLUN *et al.* 2003). We found that in all cases Ho was imported into the nucleus, although each of the zinc finger mutants showed a certain reduction in nuclear accumulation of Ho compared with native protein. This was somewhat more pronounced for the mutations in the upstream zinc fingers (Figure 6). Nuclear accumulation of these mutant Ho proteins resembles the less efficient karyopherin-mediated import of PI-*Sce*I in meiotic cells (NAGAI *et al.* 2003).

Superposition of the Ho model onto the aligned structures of I-*Cre*I, PI-*Sce*I, and I-*Dmo*I and identification of catalytic site residues: The crystal structures of the four LAGLIDADG endonucleases I-*Cre*I, PI-*Sce*I, I-*Dmo*I, and PI-*Pfu*I can be superposed on the basis of the central helices formed by the LAGLIDADG sequence. The I-*Cre*I dimer corresponds to the pseudosymmetric endonucleolytic domain of monomeric PI-*Sce*I, I-*Dmo*I, and PI-*Pfu*I. The enzymes thus contain two symmetric catalytic sites with each site cleaving one DNA strand of the cognate sequence across the minor groove (JURICA *et al.* 1998; CHEVALIER *et al.* 2001).

The superposition of structures shows that diverse residues are found in the periphery of the active sites and that only the acidic side chain of the LAGLIDADG sequence that coordinates the divalent cation in the I-*Cre*I/DNA structure (Asp20) is conserved in all four structures (CHEVALIER and STODDARD 2001). Cleavage of the phosphodiester bond requires a general base to activate the nucleophile, a Lewis acid to stabilize the pentacoordinate transition state, and a general acid to protonate the 3′ leaving group. I-*Cre*I uses a three-metal mechanism, the central metal being shared by two separate active sites (CHEVALIER *et al.* 2001). The general base and acid appear to be within a well-ordered solvent network positioned by peripheral residues and the first metal ion. The protein makes no direct contact to the nucleophile, the scissile phosphate, or the 3′ leaving group.

Guided by the superpositions and homology model we carried out mutagenesis studies on putative critical residues in the active site of Ho (Figure 2C). Asp218 and Asp326 are the metal-coordinating residues of the PI-*Sce*I active site; they are Asp222 and Asp333 in Ho. We mutated Asp333 of Ho to alanine by site-directed mutagenesis and this resulted in a loss of activity (Figures 3 and 4).

Pairwise comparisons between structures show that a glutamine residue (Gln47) is conserved between I-*Cre*I and both I-*Dmo*I active sites (Gln42, the amine group in the side chain of Asn129). Mutation of this residue in I-*Cre*I and the homologous residue in another LAGLI DADG enzyme, I-*Ceu*I (Gln93), abolishes catalytic activity (SELIGMAN *et al.* 1997; TURMEL *et al.* 1997). No homologous residue is present in either of the two PI-*Sce*I active sites. Conversely, a lysine residue is conserved between I-*Cre*I (Lys98 and 98′) and both PI-*Sce*I active sites (Lys301, Lys403), but is found at only one active site in I-*Dmo*I (Lys120; SILVA *et al.* 1999) and of PI-*Pfu*I (ICHIYANAGI *et al.* 2000; CHEVALIER and STODDARD 2001). In the I-*Cre*I/DNA structure this lysine is within 6.5 Å of the scissile phosphate. Mutation of I-*Cre*I Lys98 abolishes catalytic activity (SELIGMAN *et al.* 1997). In

TABLE 1

Primers for truncations, deletions, and point mutations of Ho

Ho-185 F	CGTGATCAGATGGGTCCAGTACTCACAGGAAATGGT
Ho-R	ATACTGATCATAGCAGATGCGCGCACCTGCGT
Ho-F	TACTTTACCATGGTTTCTGAAAACACGACTATTCT
Ho-446 R	ATACTGATCATAGCAGATGCGCGCACCTGCGT
K308A	AGGGATCTTGATGGAGAG AAG/gcc[a] CAAATCCCTGAA TTTATG (*Hae*III site)[b]
K417A	TGTCGAAGTGGTCACAAA/gcg[a]ACAAGAGAAGTTCCG CCAATTATAAAA (*Bst*UI site)[b]
R286A	AGGAAGACAAGGAATTTG AGG/gcc[a] AAAAATAAT CCATTCTGGAAA (*Msc*I site)[b]
D333A	GGCTTGATCGACTCt GAT/gca GG (*Pst*I site)[b]
HoSpeF	CAGATGAGCTCACCTGCGTTGTTACCACAAC
HoSacR	GAGAAAATGCGGCCGCCTGGGGTCTCTACCTTACG
pHYTruncR	AATTAACCCTCGCTAAAGGGAACAAAACCTGGAGCTCGTGGCGCCA CCGTGCTTCTGGTACATACTTGCAATTTATACAG

[a] AAG/gcc, gcc substituted for AAG.
[b] Both the primer listed and its complementary strand were used.

PI-*Sce*I mutation of Lys301 abolishes cleavage whereas mutation of Lys403 reduces cleavage 50-fold. This difference may be related to the asymmetry of the cognate sequence (Gimble *et al.* 1998) or to the order of strand cleavage (Ichiyanagi *et al.* 2000). Ho has both symmetrically situated lysine residues (Lys308 and Lys417) and we mutated each lysine residue separately. Each residue change created a novel diagnostic restriction site (Table 1) and positive clones were confirmed by direct sequencing. Cells expressing Ho with alanine instead of either Lys308 or Lys417 did not show cleavage of *MAT* by Southern analysis (Figure 3). K417A did not induce a mating-type switch; however, the K308A mutant showed residual mating-type interconversion activity. Quantitation by serial dilutions indicated that the K308A mutation reduced mating-type switching ~10-fold (Figure 4). In PI-*Sce*I it is the K403A (Ho K417) substitution that retains residual activity.

Another basic residue observed in the catalytic site of the I-*Cre*I/DNA structure is Arg51, a side chain also identified in mutation screens. Arg51 may have its counterparts in PI-*Sce*I with Arg231 and His343, where mutations to Ala cause a small reduction in activity (Gimble *et al.* 1998). These residues are not conserved in the Ho model; they align with Leu235 and Tyr354 of Ho, respectively. However, Arg286 aligns with Asn281 of PI-*Sce*I, a residue that contacts the DNA in the cocrystal (Moure *et al.* 2002) and is shown to be critical for activity (Hu *et al.* 1999). Indeed we found that substitution of Arg286 with alanine abolished Ho activity *in vivo*. Restoration of R286A to Arg286 with a small fragment of Ho gave a switching efficiency equal to the Ho controls (Figures 3 and 4).

DISCUSSION

Homing endonucleases recognize extremely long cognate sites despite their small size and this ensures high specificity and protects the host genome from spurious cleavage. Inteins from which the endonuclease moiety has been deleted or rendered nonfunctional are capable of protein splicing and this has led to a model for intein evolution in which an ancestral protein-splicing coding sequence was colonized by an invasive endonuclease (Derbyshire *et al.* 1997). Recently an artificial bifunctional intein was reconstructed by inserting the gene for I-*Cre*I into the ORF of the GyrA mini-intein of *Mycobacterium xenopi* (Fitzsimons Hall *et al.* 2002). Intron-encoded LAGLIDADG endonucleases such as I-*Cre*I have both DNA-binding and cleavage functions (Jurica *et al.* 1998). However, in PI-*Sce*I most of the binding energy comes from interactions with domain I (Grindl *et al.* 1998). Footprinting and DNA cross-linking experiments of PI-*Sce*I followed by the crystal structure of PI-*Sce*I bound to its cognate DNA have identified the residues that contact the target DNA. Domain II residues contact base pairs −10 to +5 whereas domain I residues contact base pairs +6 to +21 upstream of the cleavage site. Within this latter region contacts between PI-*Sce*I Arg94 of the DRR and base pairs +18 and +19 stabilize a distortion of the DNA (Moure *et al.* 2002). The cognate sites of PI-*Sce*I and of Ho can be aligned, and optimal alignment is achieved by introduction of a 2-bp gap into the *HO* sequence (Gimble and Wang 1996). Critical residues identified in the *HO* cognate sequence (Nickoloff *et al.* 1986) and in the PI-*Sce*I recognition sequence coincide in the region recognized by endonuclease domain II; those recognized by domain I and the DRR are not conserved (Gimble and Wang 1996). The DRR provides some of the uniformly distributed phosphate contacts and this may be the function of Ho DRR residue Lys99 that we find to be critical for Ho activity. We find that the domain I residues contributed by the C-terminal sequence of the Hint domain (intein motifs F and G) are totally dispensable. These residues assemble in the center of the Hint horseshoe structure in functional inteins (Ichiyanagi *et al.* 2000) and are not involved in contacts with DNA in the PI-*Sce*I-DNA cocrystal (Moure *et al.* 2002).

Furthermore, our data show that the zinc finger do-

main has an essential role consistent with *in vitro* studies demonstrating that zinc ions are essential for Ho activity (Jin *et al.* 1997). Our finding that the downstream His-X2-Cys pair of the third putative zinc finger can be truncated without any effect on Ho activity implies that the successive pairs of Cys-X2-Cys sequences do not form linearly ordered classical zinc fingers, but that this region probably forms a more complex zinc finger fold with the ligands holding the zinc interspersed in the sequence (Schwabe and Klug 1994). The cognate sites of PI-*Sce*I and Ho endonuclease are very similar and both the *VMA1* gene that hosts PI-*Sce*I and *HO* are 80,000 bp apart on chromosome IV (Gimble and Wang 1996). The zinc finger domain may have been acquired during invasion of an ectopic site on this chromosome by an ancestral variant with a relaxed site specificity. Single point mutations of the *MAT* target sequence (*MAT*inc) protect the cells from Ho cleavage (Weiffenbach *et al.* 1983). The extra domains of the intein-encoded homing endonucleases, the DRR of PI-*Sce*I, the stirrup of PI-*Pfu*I, and the DRR and unique zinc finger domain of Ho may enhance the site specificity of binding, although we cannot exclude other roles. For example, the single zinc finger of intron endonuclease, I-Tev1, serves as a distance determinant in the flexible linker and determines the spacing between the site of DNA binding and the site of cleavage by the catalytic moiety (Dean *et al.* 2002).

Superposition of the structural model of Ho onto the solved structures of four LAGLIDADG endonucleases enabled us to pinpoint residues with a potential role in catalysis, to mutate them individually, and to check their role using a simple mating-type switch bioassay. The mutagenesis study shows that the active site residues in Ho are similar to those of I-*Cre*I and PI-*Sce*I. The finding that K308 rather than K417 shows residual activity, rather than the opposite as observed for PI-*Sce*I, may be due to differences between them in the architecture of the hydration sphere that coordinates the metal ions. However, despite these similarities there may be striking differences in the protein/DNA contacts as shown for the isoschizomers I-*Cre*I and I-*Mso*I (Chevalier *et al.* 2003).

The I-*Cre*I cognate site can be cleaved by four different subfamilies of the 23S rRNA intron-encoded I-*Cre*I endonuclease. These proteins share a related structure although only one-third of the residues that interact with DNA are conserved. This high sequence variability may permit the spread of homing endonucleases to new genomic locations by enabling cleavage of variant DNA sequences (Lucas *et al.* 2001). In contrast, the *HO* genes of *S. cerevisiae* and *S. bayanus* show 93% identity with 100% conservation of the catalytic residues and those we identified by mutation in domain I and in the zinc finger domain. Both genes are also extremely highly conserved (>99% identity) in the allopolyploid *S. pastorianus*, a hybrid between *S. cerevisiae* and *S.bayanus*. Each of the alleles isolated from *S. pastorianus* recognizes the *MAT* cognate site and induces a mating-type switch in *S. cerevisiae* (Tamai *et al.* 2000). The difference in sequence diversity between subfamilies of I-*Cre*I and Ho may explain why *HO* is restricted to the genus Saccharomyces.

We thank the International Union Against Cancer (UICC) for supporting a visit of D.R. to the Stoddard laboratory and the Israel Cancer Association and the Israel Cancer Research Fund (New York) for research support to D.R.

LITERATURE CITED

Belfort, M., and R. J. Roberts, 1997 Homing endonucleases: keeping the house in order. Nucleic Acids Res. **25:** 3379–3388.

Chevalier, B. S., and B. L. Stoddard, 2001 Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. **29:** 3757–3774.

Chevalier, B. S., R. J. Monnat, Jr. and B. L. Stoddard, 2001 The homing endonuclease I-*Cre*I uses three metals, one of which is shared between the two active sites. Nat. Struct. Biol. **8:** 312–316.

Chevalier, B., M. Turmel, C. Lemieux, R. J. Monnat and B. L. Stoddard, 2003 Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. J. Mol. Biol. **329:** 253–269.

Dean, A. B., M. J. Stanger, J. T. Dansereau, P. Van Roey, V. Derbyshire *et al.*, 2002 Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. Proc. Natl. Acad. Sci. USA **99:** 8554–8561.

Derbyshire, V., D. W. Wood, W. Wu, J. T. Dansereau, J. Z. Dalgaard *et al.*, 1997 Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. Proc. Natl. Acad. Sci. USA **94:** 11466–11471.

Duan, X., F. S. Gimble and F. A. Quiocho, 1997 Crystal structure of PI-*Sce*I, a homing endonuclease with protein splicing activity. Cell **89:** 555–564.

Fitzsimons Hall, M., C. J. Noren, F. B. Perler and I. Schildkraut, 2002 Creation of an artificial bifunctional intein by grafting a homing endonuclease into a mini-intein. J. Mol. Biol. **323:** 173–179.

Gimble, F. S., 2000 Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol. Lett. **185:** 99–107.

Gimble, F. S., and J. Wang, 1996 Substrate recognition and induced DNA distortion by the PI-*Sce*I endonuclease, an enzyme generated by protein splicing. J. Mol. Biol. **263:** 163–180.

Gimble, F. S., X. Duan, D. Hu and F. A. Quiocho, 1998 Identification of Lys-403 in the PI-*Sce*I homing endonuclease as part of a symmetric catalytic center. J. Biol. Chem. **273:** 30524–30529.

Grindl, W., W. Wende, V. Pingoud and A. Pingoud, 1998 The protein splicing domain of the homing endonuclease PI-*Sce*I is responsible for specific DNA binding. Nucleic Acids Res. **26:** 1857–1862.

Hall, T. M., J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy *et al.*, 1997 Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. Cell **91:** 85–97.

He, Z., M. Crist, H. Yen, X. Duan, F. A. Quiocho *et al.*, 1998 Amino acid residues in both the protein splicing and endonuclease domains of the PI-*Sce*I intein mediate DNA binding. J. Biol. Chem. **273:** 4607–4615.

Heath, P. J., K. M. Stephens, R. J. Monnat, Jr. and B. L. Stoddard, 1997 The structure of I-*Cre*I, a group I intron-encoded homing endonuclease. Nat. Struct. Biol. **4:** 468–476.

Hirata, R., Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki *et al.*, 1990 Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:** 6726–6733.

Hu, D., M. Crist, X. Duan and F. S. Gimble, 1999 Mapping of a DNA binding region of the PI-sceI homing endonuclease by affinity cleavage and alanine-scanning mutagenesis. Biochemistry **38:** 12621–12628.

Hu, D., M. Crist, X. Duan, F. A. Quiocho and F. S. Gimble, 2000 Probing the structure of the PI-*Sce*I-DNA complex by affinity

cleavage and affinity photocross-linking. J. Biol. Chem. **275:** 2705–2712.

Ichiyanagi, K., Y. Ishino, M. Ariyoshi, K. Komori and K. Morikawa, 2000 Crystal structure of an archaeal intein-encoded homing endonuclease PI-*Pfu*I. J. Mol. Biol. **300:** 889–901.

Jin, Y., G. Binkowski, L. D. Simon and D. Norris, 1997 Ho endonuclease cleaves *MAT* DNA *in vitro* by an inefficient stoichiometric reaction mechanism. J. Biol. Chem. **272:** 7352–7359.

Jurica, M. S., and B. L. Stoddard, 1999 Homing endonucleases: structure, function and evolution. Cell Mol. Life Sci. **55:** 1304–1326.

Jurica, M. S., R. J. Monnat, Jr. and B. L. Stoddard, 1998 DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-*Cre*I. Mol. Cell **2:** 469–476.

Kaplun, L., Y. Ivantsiv, D. Kornitzer and D. Raveh, 2000 Functions of the DNA damage response pathway target Ho endonuclease of yeast for degradation via the ubiquitin-26S proteasome system. Proc. Natl. Acad. Sci. USA **97:** 10077–10082.

Kaplun, L., Y. Ivantsiv, A. Bakhrat and D. Raveh, 2003 DNA damage response—mediated degradation of Ho endonuclease via the ubiquitin system involves its nuclear export. J. Biol. Chem. **278:** 48727–48734.

Klug, A., and J. W. Schwabe, 1995 Protein motifs 5. Zinc fingers. FASEB J. **9:** 597–604.

Kostriken, R., and F. Heffron, 1984 The product of the *HO* gene is a nuclease: purification and characterization of the enzyme. Cold Spring Harbor Symp. Quant. Biol. **49:** 89–97.

Kowalski, J. C., and V. Derbyshire, 2002 Characterization of homing endonucleases. Methods **28:** 365–373.

Lucas, P., C. Otis, J. P. Mercier, M. Turmel and C. Lemieux, 2001 Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. Nucleic Acids Res. **29:** 960–969.

Lykke-Andersen, J., R. A. Garrett and J. Kjems, 1996 Protein footprinting approach to mapping DNA binding sites of two archaeal homing enzymes: evidence for a two-domain protein structure. Nucleic Acids Res. **24:** 3982–3989.

Maniatis, T., E. Fritsch and J. Sambrook, 1982 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moure, C. M., F. S. Gimble and F. A. Quiocho, 2002 Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nat. Struct. Biol. **9:** 764–770.

Nagai, Y., S. Nogami, F. Kumagai-Sano and Y. Ohya, 2003 Karyopherin-mediated nuclear import of the homing endonuclease VMA1-derived endonuclease is required for self-propagation of the coding region. Mol. Cell. Biol. **23:** 1726–1736.

Nahon, E., and D. Raveh, 1998 Targeting a truncated Ho-endonuclease of yeast to novel DNA sites with foreign zinc fingers. Nucleic Acids Res. **26:** 1233–1239.

Neff, N., 1993 Protein splicing: selfish genes invade cellular proteins. Curr. Opin. Cell Biol. **5:** 971–976.

Nickoloff, J. A., E. Y. Chen and F. Heffron, 1986 A 24-bp sequence from MAT locus stimulates intergenic recombination in yeast. Proc. Natl. Acad. Sci. USA **83:** 7831–7835.

Perler, F. B., 2000 InBase, the intein database. Nucleic Acids Res. **28:** 344–345.

Perler, F. B., M. Q. Xu and H. Paulus, 1997 Protein splicing and autoproteolysis mechanisms. Curr. Opin. Chem. Biol. **1:** 292–299.

Pietrokovski, S., 1994 Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci. **3:** 2340–2350.

Russell, D., R. Jensen, M. Zoller, J. Burke, B. Errede *et al.*, 1986 Structure of the *Saccharomyces cerevisiae HO* gene and analysis of its upstream regulatory region. Mol. Cell. Biol. **6:** 4281–4294.

Schwabe, J. W., and A. Klug, 1994 Zinc mining for protein domains. Nat. Struct. Biol. **1:** 345–349.

Seligman, L. M., K. M. Stephens, J. H. Savage and R. J. Monnat, Jr., 1997 Genetic analysis of the *Chlamydomonas reinhardtii* I-*Cre*I mobile intron homing system in *E. coli.* Genetics **147:** 1653–1664.

Sherman, F., G. Fink and J. Hicks, 1986 *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Silva, G. H., J. Z. Dalgaard, M. Belfort and P. Van Roey, 1999 Crystal structure of the thermostable archaeal intron-encoded endonuclease I-*Dmo*I. J. Mol. Biol. **286:** 1123–1136.

Simon, P., P. Houston and J. Broach, 2002 Directional bias during mating type switching in Saccharomyces is independent of chromosomal architecture. EMBO J. **21:** 2282–2291.

Strathern, J. N., A. J. Klar, J. B. Hicks, J. A. Abraham, J. M. Ivy *et al.*, 1982 Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the *MAT* locus. Cell **31:** 183–192.

Tamai, Y., K. Tanaka, N. Umemoto, K. Tomizuka and Y. Kaneko, 2000 Diversity of the *HO* gene encoding an endonuclease for mating-type conversion in the bottom fermenting yeast *Saccharomyces pastorianus.* Yeast **16:** 1335–1343.

Turmel, M., C. Otis, V. Cote and C. Lemieux, 1997 Evolutionarily conserved and functionally important residues in the I-*Ceu*I homing endonuclease. Nucleic Acids Res. **25:** 2610–2619.

Weiffenbach, B., D. T. Rogers, J. E. Haber, M. Zoller, D. W. Russell *et al.*, 1983 Deletions and single base pair changes in the yeast mating type locus that prevent homothallic mating type conversions. Proc. Natl. Acad. Sci. USA **80:** 3401–3405.

Communicating editor: M. D. Rose