Function and Evolution of HO and VDE Endonucleases in Fungi

James E. Haber, Kenneth H. Wolfe

1 Introduction

The site-specific HO and VDE endonucleases are unusual members of a family of so-called group I LAGLIDADG homing endonucleases that are generally implicated in the homing of intron and intein sequences. The great majority of these endonucleases are found in mitochondria and plastids of eukaryotes, but in budding yeast two members of this family apparently "escaped" into the nucleus. In each case, the endonuclease creates a site-specific double-strand break (DSB) in a target and promotes mobility of DNA sequences by homologous recombination requiring the Rad52 and Rad51 group of recombination proteins.

The HO endonuclease has been the subject of a great deal of interest, both because of its remarkable evolution and because of its great utility in the detailed analysis of DSB-mediated recombination. Unlike most homing endonucleases, the HO endonuclease does not promote its own amplification, but rather catalyzes the switching/replacement of yeast's mating-type (*MAT*) genes.

In addition, budding yeasts harbor the VDE endonuclease, which is found as an intein in the *VMA1* gene. VDE promotes its propagation from a *VMA1* gene containing the VDE intein into a *VMA1* gene lacking such a sequence. VDE is remarkable also in sharing a close sequence and presumably evolutionary relationship with HO.

J.E. Haber (e-mail: haber@brandeis.edu)
Department of Biology and Rosenstiel Center, Brandeis University, Waltham, Massachusetts 02454–9110, USA

K.H. Wolfe (e-mail: khwolfe@tcd.ie)
Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland

Nucleic Acids and Molecular Biology, Vol. 16
Marlene Belfort et al. (Eds.)
Homing Endonucleases and Inteins

2 Mating-Type Genes in *S. cerevisiae*

Many fungi use some form of homothallic mating-type switching to produce diploids from haploid meiotic segregants. In budding yeast *MAT***a**/*MAT*α diploids are created by conjugation of haploid *MAT***a** and *MAT*α cells. The *MAT***a**

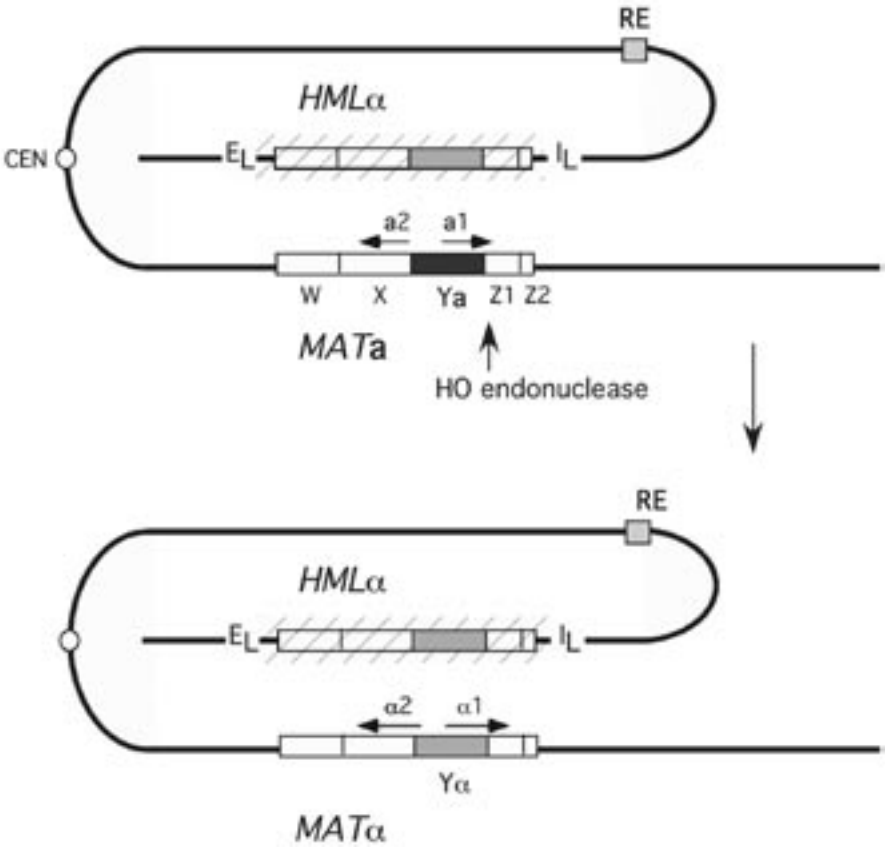

Fig. 1. Arrangement of mating-type genes in *S. cerevisiae*. The *MAT* locus is approximately 2400 bp. Mating-type-specific Y**a** and Yα sequences are about 650 and 750 bp, respectively, and are surrounded by 700-bp W and X regions, a 230-bp Z1 region and a 90-bp Z2 region. The orientation of *MAT***a** and *MAT*α transcripts are shown. The donor sequences *HML*α and *HMR***a** are unexpressed and maintained in a silent, heterochromatic state by action of the Sir2 histone deacetylase and accessory Sir1, Sir3 and Sir4 proteins, which are first recruited to E and I silencer sites surrounding the donors. *MAT* switching is initiated by HO endonuclease cleavage in Z1, cutting 7 and 3 bp on the top and bottom strands from the Y/Z1 border, creating 4-bp 3′ overhanging ends. The donor loci cannot be cut. Donor preference is governed by the recombination enhancer (RE), which controls recombination efficiency along the entire left arm of chromosome III

and *MAT*α alleles are located in the middle of chromosome III (Fig. 1) and differ from each other in a 643-bp (Y**a**) or 748-bp (Yα) region that contains different DNA sequences encoding regulators of mating type (reviewed by Haber 2002). *MAT*α encodes two genes, *MAT*α1 and *MAT*α2. *MAT*α1 encodes a positive transcriptional regulator of α-specific genes (e.g. the pheromone α-factor, and an **a**-pheromone receptor, Ste3). *MAT*α2 encodes a repressor with two modes of action. In *MAT*α cells, *MAT*α2 pairs with the general transcriptional regulator, Mcm1, to repress **a**-specific genes (e.g. those encoding the pheromone, a-factor, and the α-factor receptor, Ste2). In *MAT***a**/*MAT*α diploid cells, *MAT*α2 and *MAT***a**1 proteins form a repressor that turns off haploid-specific genes, including the pheromone signal transduction pathway, the nonhomologous end-joining (NHEJ) pathway of DNA repair and a repressor of meiosis.

Most *S. cerevisiae* strains carry two additional copies of mating-type genes, *HML*α and *HMR***a**, at distant locations on the same chromosome as the *MAT* locus (Fig. 1). In some strains the specific mating-type sequences at these loci are reversed (i.e. *HML***a** or *HMR*α). These genes are, however, not expressed, as both *HML* and *HMR* are surrounded by silencer sequences (designated E and I) that recruit the specialized histone deacetylase, Sir2, and other silencing factors (Sir1, Sir3 and Sir4) to create a heterochromatic domain that is not transcribed (Rusche et al. 2003). These silent loci are the key to the ability of HO endonuclease to promote the switching of the *MAT* locus.

3 HO-Induced *MAT* Switching in *S. cerevisiae*

Homothallic haploid cells, carrying an active HO endonuclease gene, can switch to the opposite mating-type allele. When the *MAT* locus is cleaved by HO endonuclease, the DSB is repaired by homologous recombination (gene conversion) using one of two unexpressed donor loci, *HML*α or *HMR***a**, which are located 200 and 100 kb 5′ and 3′, respectively, on chromosome III. *HML* and *HMR* differ from each other in that *HML* shares more extensive (regions W and Z2) homology with *MAT* than does *HMR*. *MAT* switching displays donor preference, in which *MAT***a** selectively recombines with *HML* whereas *MAT*α preferentially recombines with *HMR*.

Expression of HO endonuclease is strongly regulated, so that it is turned on only in the late G1 stage of the cell cycle and only in cells that have undergone at least one cell division cycle. After two cell divisions, two *MAT*α cells are juxtaposed to two *MAT***a** cells and efficient conjugation produces two *MAT***a**/MATα zygotes. Coexpression of *MAT***a** and *MAT*α turns off the HO gene itself. Thus, once cells switch, the process is complete until meiosis creates haploid segregants.

Much of the research studying HO-mediated *MAT* switching has relied on freeing the HO gene from its transcriptional controls. Jensen and Herskowitz (1984) fused the HO gene to a galactose-regulated promoter, so that all cells in the population can be induced to switch in a synchronous fashion. The use of the *GAL::HO* gene has proven to be invaluable in studying not only *MAT* switching, but in characterizing other forms of DSB repair (NHEJ and break-induced replication) as well as in the detailed study of the DSB-induced DNA damage checkpoint. The applications of HO as a regulated source of defined DSBs to study these different processes have been well reviewed (Moore and Haber 1996a,b; Toczyski et al. 1997; Lee et al. 1998; Nickoloff and Hoekstra 1998; Haber 2000; Melo and Toczyski 2002; Aylon and Kupiec 2004). Here, we focus only on *MAT* switching itself.

4 Mechanism of *MAT* Switching

The ability to induce synchronous HO cleavage at *MAT* made it possible to follow the process of *MAT* switching in "real-time" by isolating DNA samples at intervals after the creation of the DSB. A combination of Southern blot and polymerase chain reaction (PCR) analysis to analyze DNA intermediates (White and Haber 1990) as well as chromatin immunoprecipitation (ChIP) to identify recombination proteins that participate in the gene conversion process (Sugawara et al. 2003; Wolner et al. 2003; Wang and Haber 2004) have identified a number of steps in the process (here shown as *MAT***a** switching to *MAT*α in Fig. 2): (1) the resection of the DSB ends by 5′ to 3′ exonucleases to create single-strand DNA (ssDNA) ends; (2) assembly of a Rad51 filament on ssDNA to facilitate a search for homologous donor sequences to repair the DSB (i.e. *HML* or *HMR*); (3) strand invasion and the synapsis between the donor and *MAT* sequences; (4) use of the 3′ end of the invading ssDNA as a primer to promote new DNA synthesis, a process that involves PCNA and either DNA polymerases δ or ε; (5) displacement of the newly synthesized strand to pair with the second end of the DSB and the completion of the replacement of the original Y**a** sequence with Yα. During the process, the donor is left unchanged and all the newly synthesized DNA is found at the *MAT* locus (G. Ira and J.E. Haber, unpubl.).

5 Donor Preference Associated with *MAT* Switching

There is one other remarkable aspect of *MAT* switching: the mating-type regulated choice of *HML* or *HMR* as a donor in repairing the HO-induced DSB. A *MAT***a** haploid recombines with *HML*α about 90% of the time, whereas a

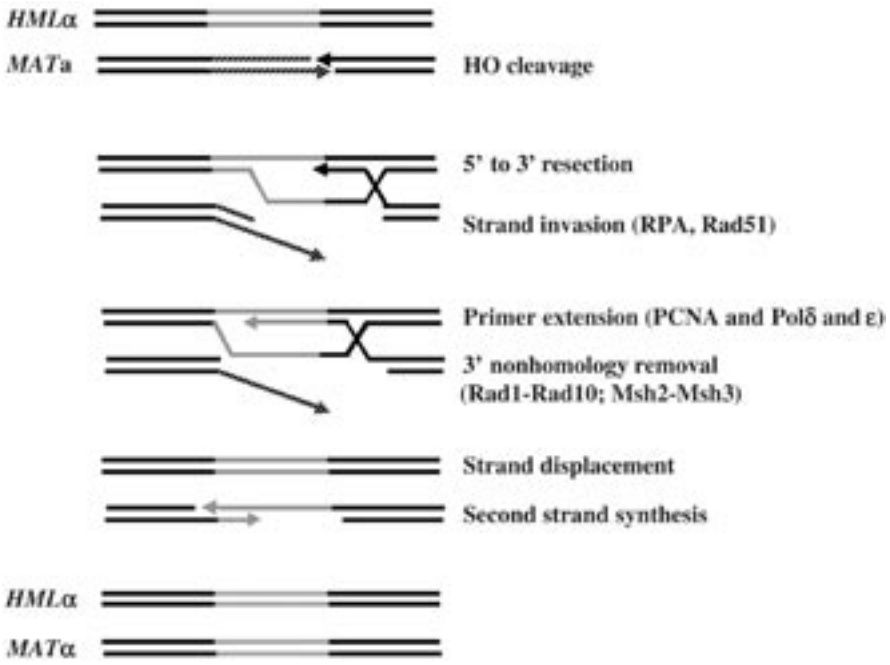

Fig. 2. Mechanism of *MAT***a** switching. Following the creation of an HO-induced DSB, the 5′ ends are resected to create 3′-ended single-stranded tails that are bound first by the single-stranded binding protein complex RPA and then by the Rad51 recombinase that promotes strand invasion of the Z region of the *HML*α donor. The 3′ end of the invading DNA is used as a primer to initiate new DNA synthesis. The newly synthesized strand is apparently displaced and pairs with resected DNA on the other side of the DSB. *MAT* switching is completed, almost always without an accompanying crossover, by copying the second strand

*MAT*α haploid selectively chooses *HMR***a**. This ensures that, most of the time, the process of switching will produce an equal number of juxtaposed cells of opposite mating type after two cell divisions. "Donor preference" is not determined by the differences between the silencer regions surrounding *HML* and *HMR*, nor by the Y**a** or Yα sequences themselves; in a strain in which *HML*α is deleted and replaced by a cloned *HMR*α locus, *MAT***a** cells continue to select the left-arm donor (Weiler and Broach 1992). Control of donor preference lies with the recombination enhancer (RE), a small, cis-acting sequence located 17 kb away from *HML* (Wu and Haber 1996; Wu et al. 1998). In *MAT***a** cells, RE is active and facilitates the use of *HML* over *HMR*. In *MAT*α cells, RE is repressed, and the left-arm donor becomes in some way inaccessible; consequently, the right-arm donor is used preferentially. RE acts over

the entire 115 kb of the left arm of chromosome III (Wu and Haber 1996; Sun et al. 2002). A donor placed anywhere along this arm can be used selectively in *MAT***a** cells. When RE is deleted, the left-arm donor becomes inaccessible in *MAT***a** cells. Most of the activity of RE is contained in a segment of about 750 bp. Further analysis, including comparisons between the REs of three *Saccharomyces* species (Wu et al. 1998) and the creation of synthetic RE sequences from multimers of conserved subregions (Wu and Haber 1996; Sun et al. 2002), showed that RE consists of several highly conserved domains, including two regions with ten or more iterations of TTT(A/G) and a highly conserved MAT "Greek alpha" 2-Mcm1 "operator region" which shares strong sequence identity with sequences that control expression of "**a**-specific genes".

Repression of RE in *MAT*α cells occurs by binding of the α2-Mcm1 repressor that also turns off **a**-specific genes to the operator site and leads to the establishment of highly positioned nucleosomes across the 2.5-kb region containing RE (there are no open reading frames in this region; there is no change in the chromatin structure of any of the genes along the chromosome arm nor changes in the silencing of *HML*). In *MAT***a**, the positioned nucleosomes are absent and several DNaseI-hypersensitive sites indicative of protein binding are found (Wu et al. 1998). Activation of RE depends on Mcm1 (which is also an activator of expression of **a**-specific genes). Mcm1 binding facilitates the binding of the Fkh1 transcription regulator (Sun et al. 2002). The complex RE sequences can be replaced by as few as four 22-bp copies of one of the Fkh1-binding sites located in several conserved subregions of the RE. Small effects on *MATa*'s use of *HML* have been seen when the cohesin assembly protein Chl1 is deleted (Weiler et al. 1995).

How RE acts remains unknown, but experiments recruiting the LacI-GFP fusion protein to LacO arrays inserted near *HML*, *MAT* or *HMR* have provided some clues. Live-cell imaging of GFP-tagged chromosomes suggests that the left arm of chromosome III is more confined in *MAT*α (or RE-deleted) cells than in *MAT*a (Bressan et al. 2004). The molecular basis of this tethering remains under investigation.

6 Evolutionary Origins of the *HO* Gene and Other Components of the *MAT* Switching System

The evolutionary origin of *HO* is enigmatic because the gene has a very limited phylogenetic distribution. Apart from *S. cerevisiae* and its close relatives such as *S. bayanus* (the *Saccharomyces sensu stricto* group of species), the *HO* gene is present only in *Candida glabrata, Saccharomyces castellii,* and *Zygosaccharomyces rouxii* (Butler et al. 2004). In these species, mating-type switching probably occurs by HO-catalyzed switching between an active *MAT*

locus and silent *HM* cassettes, similar to the *S. cerevisiae* paradigm, although a direct demonstration of such switching has yet to be shown. A second group of yeasts, exemplified by *Kluyveromyces lactis*, have *HM* cassettes but do not have an ortholog of *HO*. *K. lactis* can switch mating types at a very low frequency (Herman and Roman 1966; Zonneveld and Steensma 2003), similar to HO-negative *ho* strains of *S. cerevisiae*. Other species with cassettes but lacking *HO* include *K. waltii* and *A. gossypii*. Switching in these species probably proceeds by homologous recombination between the cassettes and the *MAT* locus, without a specific endonuclease to initiate the recombination process. A third group of more distantly related species, such as *Yarrowia lipolytica*, do not have silent cassettes or an *HO* gene and do not switch mating type; they are heterothallic. The *HO* gene seems to have been gained by yeasts shortly before the whole genome duplication (WGD) that occurred in the shared ancestor of several yeast species (Wolfe and Shields 1997; Dietrich et al. 2004; Kellis et al. 2004), because all known species that are descended from the WGD event have *HO* (Butler et al. 2004). The origin of *HO* can be inferred to pre-date the WGD because *HO* is present in *Z. rouxii*, which is a representative of the lineage of yeasts that is sister to the WGD lineage (Kurtzman and Robnett 2003). Gene order at the *HO* locus in different yeast species is well conserved once the WGD event is taken into account (Butler et al. 2004), and does not provide much in the way of clues to *HO*'s origins. The gene simply seems to materialize in the interval between the genes *RIO2* and *SSB*. This interval seems to have had an unusually dynamic history but the significance of this observation is unclear. In several species the interval contains a very rapidly evolving gene (*YNL208W*) with prion-like amino acid composition, and in *Z. rouxii* a gene for an endoribonuclease of the T_2 family (distantly related to *S. cerevisiae RNY1*; MacIntosh et al. 2001) is present.

7 Linkage of an *HM* Cassette with the Recombination Enhancer (RE)

In *C. glabrata* one *HM* cassette is near a telomere of the same chromosome as the *MAT* locus (chromosome 2) but the other is on chromosome 5, also near a telomere (Dujon et al. 2004). The chromosome 2 interval between *HM* and *MAT* spans the region (*KAR4–SPB1*) where the RE is located in *S. cerevisiae*, but the RE sequence motifs themselves are apparently not conserved outside the *Saccharomyces sensu stricto* species (Zhou et al. 2001). The most conserved element within the RE sequence – a binding site for the α2-Mcm1 repressor – is notably absent in the more distantly related (*sensu lato*) species. The stretch of DNA spanning *MAT*, the putative RE, and an *HM* cassette is one of the largest unrearranged genomic regions shared by *K. lactis* and

K. waltii, and is also essentially unrearranged in *A. gossypii*, which suggests that the chromosomal structure has been preserved by natural selection. Much of the gene order in this interval is also shared with *S. cerevisiae* and *C. glabrata*, but there have been some additional rearrangements in those species as well as the effect of WGD.

It should be noted that in an efficiently switching species such as *S. cerevisiae*, RE actually does not seem very important, as random use of *HML*α and *HMR***a** would yield predominantly **a**/α diploids in a few generations. But, at least in *S. cerevisiae*, the absence of RE does not give random switching; instead both *MAT***a** and *MAT*α cells use *HMR* 90% of the time (Wu and Haber 1996). Whether this preferential "default" use of one donor is yet another aspect of the evolution of chromosome III is not yet clear. If a donor on the same chromosome arm is much more likely to be used in repairing a DSB and if HO cleavage is less efficient in other species, RE would continue to be an important feature of homothallic switching.

8 The HO Endonuclease Site in *MAT*α1

The origin of HO is intimately associated with the structure of the *MAT* locus. In *S. cerevisiae*, the **a** and α idiomorphs (versions) of the *MAT* locus differ only by their Y**a** or Yα regions. Y**a** contains the complete coding region of the **a**-specific *MAT***a**1 gene, but Yα contains only the 5′ portions of the α-specific genes *MAT*α1 and *MAT*α2 which are transcribed divergently (Fig. 1). This arrangement ensures that each of the three *MAT* genes is only expressed in the correct cell type, but it also means that the 3′ ends of the *MAT*α1 and *MAT*α2 genes are located in DNA common to both idiomorphs. Thus, the junctions between unique (Y) and flanking (X or Z1) DNA occur within the coding regions of the *MAT*α1 and *MAT*α2 genes (Fig. 1). The HO cleavage site abuts the Y/Z1 junction. Cleavage of *MAT* by HO has not been experimentally demonstrated in species other than *S. cerevisiae*, but, in every species that has an *HO* gene, the Y/Z1 junction occurs at a conserved point in *MAT*α1 which corresponds to the HO cleavage site in *S. cerevisiae*. This arrangement strongly suggests that mating-type switching in those species occurs by a mechanism analogous to the *S. cerevisiae* mechanism. In contrast, in species that do not have an *HO* gene, such as *K. lactis* and *S. kluyveri*, the Y/Z1 junction is downstream of the stop codon of *MAT*α1 and is not at a conserved position (Butler et al. 2004). Thus, the co-option of HO into the switching process seems to have "locked" the Y/Z1 junction to a particular site in *MAT*α1, whereas previously its location could drift.

Recent sequence data from *Saccharomyces sensu stricto* species have revealed that the X and Z1 sequences that facilitate mating-type switching

have been extraordinarily well conserved during evolution. Except for a few small insertions and deletions, there is 100% DNA sequence conservation among the five species *S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii* and *S. bayanus* across the entire X region (703 bp) and 150 bp of the Z1 region, whereas interspecies comparisons of Yα or Y**a** regions show much more divergence (79–89% identity among species in Yα, and 84–93% in Y**a**). This phenomenon was first noted by Kellis et al. (2003) who reported that the "MAT**a**2" sequence was completely conserved. ("*MAT***a**2" is a name given to the duplicated copy of the 3′ end of *MAT*α2 which is present in the X region and so is present in *MAT***a** idiomorphs; it is transcribed but there is no evidence that it encodes a functional protein.) However, the 100% conservation extends into the noncoding DNA downstream of *MAT*α2 and *MAT*α1. The X and Z1 regions are among the largest completely conserved DNA sequences in the genomes of these yeasts. Their apparent intolerance of nucleotide substitutions must be due to strong selection for efficient mating-type switching, together with the fact that the X and Z1 sequences each occur in three copies in the genome (at *MAT, HML* and *HMR*) which makes coevolution difficult.

9 Relationship of HO to VDE

The HO protein has strong sequence similarity to inteins, which are unusual selfish genetic elements found primarily in bacteria (see Perler, this Vol.). Only two nuclear genes of eukaryotes are known to contain inteins (Perler 2002), one of which is the vacuolar H^+-ATPase gene *VMA1* of *S. cerevisiae* which contains the VDE intein (Gimble and Thorner 1992). HO has higher sequence similarity to VDE than to any other intein and clusters with it in a phylogenetic tree (Dalgaard et al. 1997; Gogarten et al. 2002). This fact, together with their co-occurrence in *S. cerevisiae* despite the rarity of inteins in eukaryotes, suggests that HO shares a relatively recent common ancestor with VDE and is a sort of renegade intein. Although the *VMA1* ATPase gene is a highly conserved gene with homologues in all eukaryotes, eubacteria and archaea, the distribution of the VDE intein is limited to a few hemiascomycete species closely related to *S. cerevisiae*. Among these species, VDE has a patchy phylogenetic distribution (e.g., it is present in *Candida tropicalis* but not *C. albicans*) and there is evidence that it has been horizontally transferred among yeasts (Koufopanou et al. 2002; Okuda et al. 2003).

Now that the sequences of several HO and VDE proteins are known from different species it is possible to examine patterns of domain conservation, within HO and VDE separately, using the crystal structure of the *S. cerevisiae* VDE protein (also called PI-SceI) to help interpretation (Moure et al. 2002; Bakhrat et al. 2004). Koufopanou et al. (2002) pointed out that mobile inteins

are under continual selection for efficient protein splicing to maintain expression of the host gene, but their homing endonuclease activity is only "tested" by natural selection if there are inteinless alleles available for colonization. Thus, once an intein-containing allele has reached a 100% frequency in a population its endonuclease activity can decay. It is still necessary to retain an open reading frame in the endonuclease region of the gene so that the downstream parts of the protein-splicing domain and the host gene are translated, but there will be no selection to preserve amino acid residues needed for the nuclease activity. Indeed, only three of 13 VDE proteins from different yeast species tested by Posey et al. (2004) were active endonucleases. When HO and VDE proteins are compared among a uniform set of four yeast species it is apparent that the endonuclease domain of HO is much better conserved than that of VDE (Fig. 3). However, it should be noted that because VDE can be horizontally transferred among species, it is possible that the amount of evolutionary time involved in these comparisons is not equal for the two proteins. Of the VDE proteins in this alignment, only the *S. cerevisiae* one has been shown to have an active endonuclease; the *Z. rouxii* VDE has no nuclease activity (Posey et al. 2004), and the other two have not been tested.

HO also shows strong conservation of the N-terminal part of the protein-splicing domain, even though *HO* is a free-standing gene, and there is no host gene whose product needs to be spliced. Furthermore, deletion of the N-terminal 112 residues of HO, which correspond to the protein-splicing domain, results in loss of endonuclease activity (Bakhrat et al. 2004). One possible explanation for this conservation is that the protein-splicing domain (properly called domain I) of HO may contain some residues that bind to DNA, similar to the DRR (DNA recognition region) residues in domain I of VDE (Moure et al. 2002). However, the DRR residues are poorly conserved among VDE sequences, let alone HO sequences, so it is impossible to say whether DNA recognition is the sole reason for the conservation of the protein-splicing domain in HO. It is striking that this domain of VDE, which supports protein splicing, has tolerated more insertions and deletions than the equivalent region of HO, which does not (Fig. 3). The C-terminus of HO contains a putative zinc finger domain (Russell et al. 1986), which is an extension relative to VDE. Analysis of the *S. cerevisiae* sequence alone suggested that the domain has three fingers, each with two Cys/Cys or Cys/His pairs (Russell et al. 1986; Bakhrat et al. 2004). However, the last pair (H-X2-C at positions 574–577 of the *S. cerevisiae* protein) is not conserved in other species, and the conserved Cys residues in the first pair have a C-X_3-C rather than the usual C-X_2-C spacing, making it unlikely that this pair is actually part of a zinc finger (Fig. 3). Thus the original proposals for the structure of the zinc finger in HO may not be correct.

One important difference between VDE and HO is that VDE cleavage and homing occurs only during sporulation of diploids, whereas HO is only active

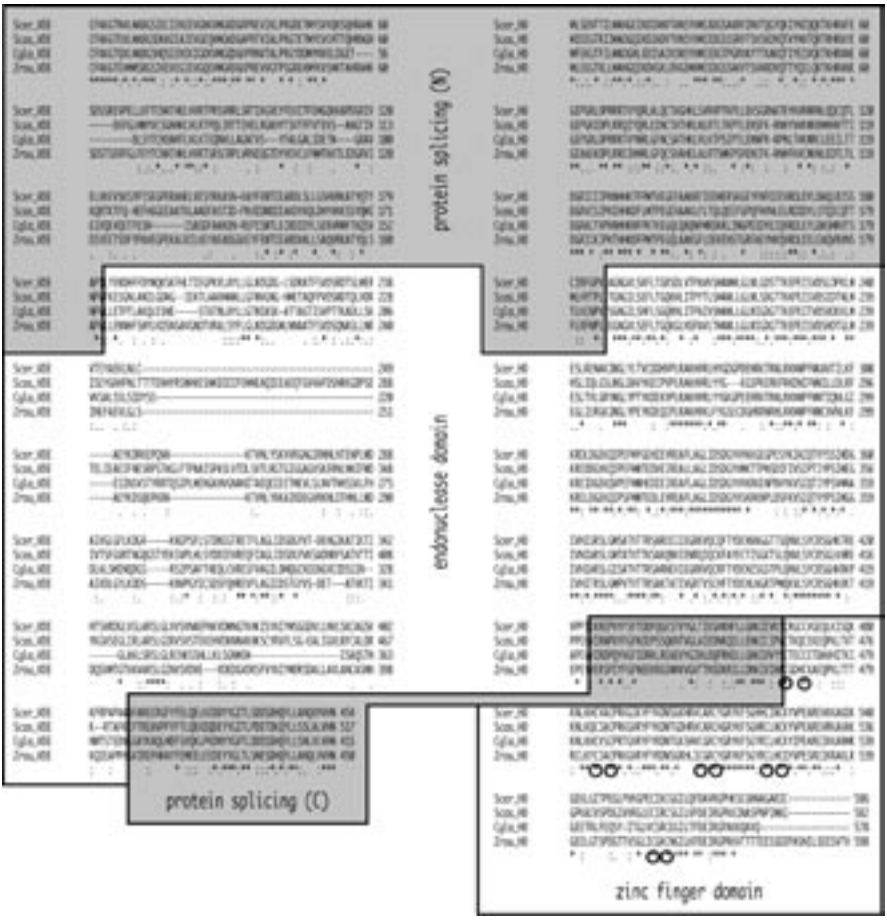

Fig. 3. ClustalW alignments of VDE proteins (*left*) and HO proteins (*right*) compared among *S. cerevisiae, S. castellii, C. glabrata* and *Z. rouxii*. The endonuclease domain of each protein lies between the N- and C-terminal parts of the protein-splicing domain, which are *shaded*. *Circles* mark conserved cysteine residues in the zinc finger of HO

in the late mitotic G1 phase of haploids. The complex way in which HO activity is regulated is reasonably well understood (Cosma 2004), whereas how VDE activity is restricted to meiosis – even though the protein is expressed at high levels in mitotically grown haploid or diploid cells – is unknown (Gimble and Thorner 1992).

10 Hypothesis for the Evolutionary Origin of HO

Where did *HO* come from? It seems very likely that HO and VDE share a recent common ancestor in the hemiascomycetes because their sequences and phylogenetic distributions are similar (Gimble and Thorner 1992; Keeling and Roger 1995; Liu 2000). One scenario is that a mobile intein (presumably from a bacterium) first invaded *VMA1* to form VDE, and has subsequently been horizontally transmitted with frequent gains and losses in various hemiascomycete species to give its current phylogenetic distribution (Koufopanou et al. 2002). The phylogenetic range over which VDE can spread is limited by the requirement for a suitable target site in the *VMA1* sequence of the host species, but this target site can co-evolve with the protein sequence of VDE (Posey et al. 2004). At some point in the cycle of horizontal transfers, gains and losses, the VDE of one species duplicated to give rise to a homing endonuclease gene that was the ancestor of HO. An alternative scenario is that the first mobile intein to invade the hemiascomycetes was located in some other unknown gene, and both VDE and HO genes were derived from it by later duplication.

We speculate that *HO* may have been formed when a mobile intein invaded a gene coding for a protein with a zinc finger domain. The integration site was either very close to the 5′ end of the gene, or else the original 5′ end of the zinc finger protein gene was lost later. Loss of protein-splicing activity led to the formation of a chimeric protein with an endonuclease domain and two DNA-binding regions: one in the zinc finger and one in the N-terminal protein-splicing domain. Somehow this fusion protein began to bind DNA specifically at a site in *MAT*α1, rather than in its own gene, thus creating a DSB that was repaired using the silent cassettes of mating-type information. This step may not be quite as far-fetched as it seems, because VDE recognition sites have been shown to drift (Posey et al. 2004), and there is a small amount of DNA sequence similarity between the recognition sites of HO and *S. cerevisiae* VDE (Gimble and Wang 1996; Bakhrat et al. 2004).

The origin of *HO* may also be connected to the origin of an unlinked gene, *SIR1*. A crucial feature of HO activity is that it cleaves the recognition site present in the active *MAT* locus, but not the sites in the *HM* cassettes which have identical DNA sequences. Cleavage at the silent cassettes is repressed by the presence of heterochromatin, whose formation is directed by the Sir proteins. Whereas Sir2, Sir3 and Sir4 proteins are also involved in transcriptional silencing at rDNA and telomeres, Sir1 is uniquely involved in silencing at the *HM* loci, by recruiting the NAD-dependent histone deacetylase Sir2 (Chien et al. 1993). *SIR1* genes have almost exactly the same phylogenetic distribution as HO genes: present in *Saccharomyces sensu stricto, Saccharomyces castellii* and *Zygosaccharomyces rouxii*, but absent in

K. lactis, K. waltii, S. kluyveri, A. gossypii and more distantly related species. The only discrepancy is that *C. glabrata* has *HO* but not *SIR1*.

The co-option of HO by yeasts led to a dramatic increase in the rate of mating-type switching, from once per million cells to almost one switch per cell generation, resulting in a change from a life cycle where the major growth phase was haploid (like *K. lactis* today) to one where most cells are diploid because budding of a haploid cell is followed almost inevitably by mother–daughter mating. This change was made possible by the fortuitous invasion of a parasitic genetic element but had a profound effect on the biology of a large clade of yeast species.

Acknowledgements. We thank Bernard Dujon, Cécile Fairhead and Gilles Fischer for stimulating conversations. Research in K.H.W.'s laboratory is supported by Science Foundation Ireland. Research in J.E.H.'s laboratory is supported by the US National Institutes of Health.

References

Aylon Y, Kupiec M (2004) DSB repair: the yeast paradigm. DNA Repair (Amst) 3:797–815

Bakhrat A, Jurica MS, Stoddard BL, Raveh D (2004) Homology modeling and mutational analysis of Ho endonuclease of yeast. Genetics 166:721–728

Bressan DA, Vazquez J, Haber JE (2004) Mating type-dependent constraints on the mobility of the left arm of yeast chromosome III. J Cell Biol 164:361–371

Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH (2004) Evolution of the *MAT* locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci USA 101:1632–1637

Chien CT, Buck S, Sternglanz R, Shore D (1993) Targeting of SIR1 protein establishes transcriptional silencing at HM loci and telomeres in yeast. Cell 75:531-541

Cosma MP (2004) Daughter-specific repression of *Saccharomyces cerevisiae* HO: Ash1 is the commander. EMBO Rep 5:953–957

Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS (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

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304:304–307

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J-M, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confanioleri F, de Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud J.-M, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard G.-F, Straub M-L, Suleau A, Swennen D, Tekaia

F, Wésolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet J-L (2004) Genome evolution in yeasts. Nature 430:35–44

Gimble FS, Thorner J (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. Nature 357:301–306

Gimble FS, 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

Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E (2002) Inteins: structure, function, and evolution. Annu Rev Microbiol 56:263–287

Haber JE (2000) Lucky breaks: analysis of recombination in *Saccharomyces*. Mutat Res 451:53–69

Haber JE (2002) Switching of *Saccharomyces cerevisiae* mating-type genes. In: Craig RCN, Gellert M, Lambowitz A (eds) Mobile DNA II. ASM Press, Washington, DC, pp 927–952

Herman A, Roman H (1966) Allele specific determinants of homothallism in *Saccharomyces lactis*. Genetics 53:727–740

Jensen RE, Herskowitz I (1984) Directionality and regulation of cassette substitution in yeast. Cold Spring Harbor Symp Quant Biol 49:97–104

Keeling PJ, Roger AJ (1995) The selfish pursuit of sex. Nature 375:283

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617–624

Koufopanou V, Goddard MR, Burt A (2002) Adaptation for horizontal transfer in a homing endonuclease. Mol Biol Evol 19:239–246

Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417–432

Lee SE, Moore JK, Holmes A, Umezu K, Kolodner R, Haber JE (1998) *Saccharomyces* Ku70, Mre11/Rad50 and RPA proteins regulate adaptation to G2/M arrest DNA damage. Cell 94:399–409

Liu XQ (2000) Protein-splicing intein: genetic mobility, origin, and evolution. Annu Rev Genet 34:61–76

MacIntosh GC, Bariola PA, Newbigin E, Green PJ (2001) Characterization of Rny1, the *Saccharomyces cerevisiae* member of the T2 RNase family of RNases: unexpected functions for ancient enzymes? Proc Natl Acad Sci USA 98:1018–1023

Melo J, Toczyski D (2002) A unified view of the DNA-damage checkpoint. Curr Opin Cell Biol 14:237–245

Moore JK, Haber JE (1996a) Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature 383:644–646

Moore JK (1996b) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol Cell Biol 16:2164–2173

Moure CM, Gimble FS, Quiocho FA (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nat Struct Biol 9:764–770

Nickoloff JC, Hoekstra MF (1998) Double-strand break and recombinational repair in *Saccharomyces cerevisiae*. In: Hoekstra MF (ed) DNA damage and repair, vol 1. DNA repair in prokaryotes and lower eukaryotes. Humana Press, Totowa, NJ

Okuda Y, Sasaki D, Nogami S, Kaneko Y, Ohya Y, Anraku Y (2003) Occurrence, horizontal transfer and degeneration of VDE intein family in *Saccharomycete* yeasts. Yeast 20:563–573

Perler FB (2002) InBase: the intein database. Nucleic Acids Res 30:383–384
Posey KL, Koufopanou V, Burt A, Gimble FS (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. Nucleic Acids Res 32:3947–3956
Rusche LN, Kirchmaier AL, Rine J (2003) The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. Annu Rev Biochem 72:481–516
Russell DW, Jensen R, Zoller MJ, Burke J, Errede B, Smith M, Herskowitz I (1986) Structure of the *Saccharomyces cerevisiae* HO gene and analysis of its upstream regulatory region. Mol Cell Biol 6:4281–4294
Sugawara N, Wang X, Haber JE (2003) In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. Mol Cell 12:209–219
Sun K, Coic E, Zhou Z, Durrens P, Haber JE (2002) Saccharomyces forkhead protein Fkh1 regulates donor preference during mating-type switching through the recombination enhancer. Genes Dev 16:2085–2096
Toczyski DP, Galgoczy DJ, Hartwell LH (1997) CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. Cell 90:1097–1106
Wang X, Haber JE (2004) Role of *Saccharomyces* single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. PLoS Biol 2:104–111
Weiler KS, Broach JR (1992) Donor locus selection during *Saccharomyces cerevisiae* mating type interconversion responds to distant regulatory signals. Genetics 132:929–942
Weiler KS, Szeto L, Broach JR (1995) Mutations affecting donor preference during mating type interconversion in *Saccharomyces cerevisiae*. Genetics 139:1495–1510
White CI, Haber JE (1990) Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*. EMBO J 9:663–673
Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713
Wolner B, van Komen S, Sung P, Peterson CL (2003) Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. Mol Cell 12:221–232
Wu C, Weiss K, Yang C, Harris MA, Tye BK, Newlon CS, Simpson RT, Haber JE (1998) Mcm1 regulates donor preference controlled by the recombination enhancer in *Saccharomyces* mating-type switching. Genes Dev 12:1726–1737
Wu X, Haber JE (1996) A 700 bp cis-acting region controls mating-type dependent recombination along the entire left arm of yeast chromosome III. Cell 87:277–285
Zhou Z, Sun K, Lipstein EA, Haber JE (2001) A *Saccharomyces servazzii* clone homologous to *Saccharomyces cerevisiae* chromosome III spanning KAR4, ARS 304 and SPB1 lacks the recombination enhancer but contains an unknown ORF. Yeast 18:789–795
Zonneveld BJM, Steensma HY (2003) Mating, sporulation and tetrad analysis in *Kluyveromyces* lactis. In: Barth G (ed) Non-conventional yeasts in genetics, biochemistry and biotechnology. Springer, Berlin Heidelberg New York, pp 151–154