Nucleic Acids Research, 1997, Vol. 25, No. 17 **3379–3388**

Homing endonucleases: keeping the house in order

Marlene Belfort* and Richard J. Roberts[1]

Molecular Genetics Program, Wadsworth Center, New York State Department of Health, and School of Public Health, State University of New York at Albany, PO Box 22002, Albany, New York 12201-2002, USA and [1]New England Biolabs, 32 Tozer Road, Beverly, MA 01915-5599, USA

Received May 27, 1997; Revised and Accepted July 18, 1997

ABSTRACT

Homing endonucleases are rare-cutting enzymes encoded by introns and inteins. They have striking structural and functional properties that distinguish them from restriction enzymes. Nomenclature conventions analogous to those for restriction enzymes have been developed for the homing endonucleases. Recent progress in understanding the structure and function of the four families of homing enzymes is reviewed. Of particular interest are the first reported structures of homing endonucleases of the LAGLIDADG family. The exploitation of the homing enzymes in genome analysis and recombination research is also summarized. Finally, the evolution of homing endonucleases is considered, both at the structure-function level and in terms of their persistence in widely divergent biological systems.

INTRODUCTION

Several endonucleases encoded by introns and inteins in the three biological kingdoms have been shown to promote the homing of their respective genetic elements into allelic intronless and inteinless sites (reviewed in 1–3). By making a site-specific double-strand break in the intronless or inteinless alleles, these nucleases create recombinogenic ends which engage in a gene conversion process that duplicates the intron or intein (Fig. 1). The homing enzymes that initiate the mobility process can be grouped into families, which share structural and functional properties with each other and with some freestanding, intergenic endonucleases. Regardless of whether these enzymes have been shown to be involved in DNA rearrangements, they are collectively termed homing endonucleases.

DISTINGUISHING CHARACTERISTICS OF HOMING ENDONUCLEASES

Although most homing endonucleases share with restriction enzymes the ability to make a site-specific double-strand break in the DNA target, they differ in structure, recognition properties, and genomic location (Table 1). First, the vast majority of homing endonucleases fall within one of four families, characterized by the sequence motifs LAGLIDADG, GIY-YIG, H-N-H and His-Cys box (1). In contrast, restriction enzymes do not fall within easily recognizable families. Although the $PDX_{9–18}$ (E/D)XK motif has been associated with the catalytic center of several restriction enzymes (4–6), its occurrence by chance is so frequent that it alone cannot be considered indicative of endonuclease function. Type II restriction enzymes only share significant sequence similarity if they are isoschizomers (recognize identical DNA sequences).

Second, homing endonucleases have recognition sequences that span 12–40 bp of DNA, whereas restriction enzymes recognize much shorter stretches of DNA, in the 3–8 bp range (7–9). Although the homing endonucleases are rather tolerant of single-base-pair changes in their lengthy DNA interaction sites, the restriction enzymes are highly sensitive to single-site mutations in their short recognition sequences (8–10). Furthermore, general asymmetry of homing endonuclease target sequences contrasts with the characteristic dyad symmetry of most restriction enzyme recognition sites (7–9).

Third, the enzymes have different molecular associations. Homing endonucleases act as monomers or homodimers and, while some function independently of accessory molecules, others require associated proteins to regulate their activity (11). Yet other homing endonucleases form ribonucleoprotein (RNP) complexes, wherein RNA molecules are integral components of the catalytic apparatus (12). Restriction enzymes can also function either alone, as monomers or homodimers (the Type II enzymes) (10), or with additional protein subunits (the Type I and Type III enzymes) (13), but the accessory subunits are quite different from those of the homing endonucleases. Thus, the Type I enzymes require restriction, modification, and specificity subunits for their action, while the Type III enzymes require only modification subunits for cleavage to occur (13).

Finally, the phylogenetic distribution of the two types of enzymes differs. Homing endonucleases have been found in all three biological kingdoms—the archaea, bacteria, and eukarya—whereas restriction enzymes occur only in archaea, bacteria and certain eukaryotic viruses (7–9). In addition to being phylogenetically widespread, homing endonucleases are expressed in different compartments of the eukaryotic cell: nuclei; mitochondria; and chloroplasts. Their genomic microenvironment also differs. The homing endonuclease open reading frames occur in introns, inteins, and in freestanding form between genes, whereas restriction enzyme genes have been found only in freestanding form, almost always in close association with genes encoding cognate modifying enzymes (14). Thus, while the restriction enzymes and homing endonucleases share the function of cleaving double-stranded DNA, they appear to have evolved independently.

*To whom correspondence should be addressed. Tel: +1 518 473 3345; Fax: +1 518 474 3181; Email: marlene.belfort@wadsworth.org

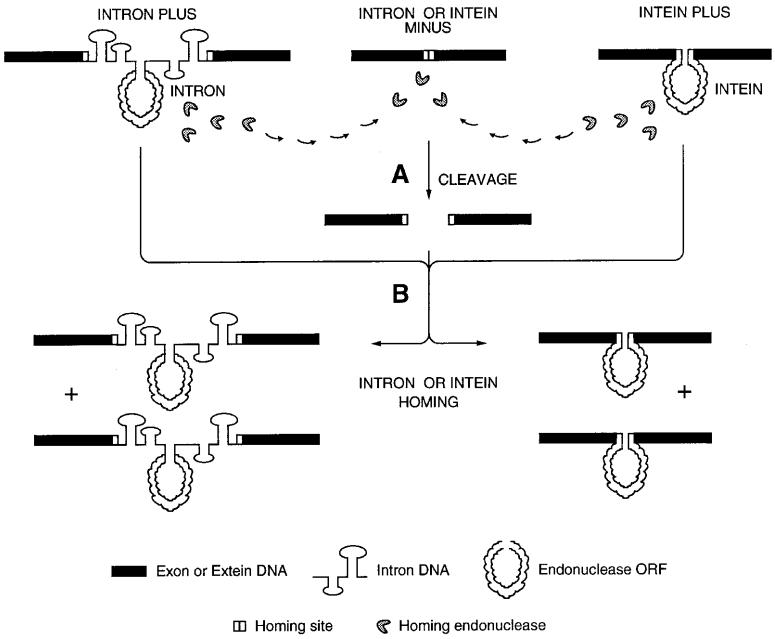

Figure 1. Intron and intein homing. (**A**) Cleavage by the homing endonuclease encoded by the intron or intein results in a double-strand break in the intron- or intein-minus allele. (**B**) The cleaved recipient engages in recombination with the intron or intein donor allele resulting in duplication of the element in a homing event.

Table 1. Comparison of homing endonucleases and restriction enzymes

Property	Homing endonuclease	Restriction enzyme
1. Conserved protein motifs	Four i. LAGLIDADG ii. GIY-YIG iii. H-N-H iv. His-Cys	None definitive[1]
2. Recognition sequences	a. Lengthy (12–40 bp) b. Asymmetric c. Sequence-tolerant	a. Short (3–8 bp) b. Symmetric and asymmetric c. Sequence-specific
3. Accessory molecules	Some require protein or RNA components for full activity	Some require methyltransferase or specificity subunits
4. Genomic location	a. Intron, intein, or intergenic b. All three biological kingdoms	a. Flanking modification gene b. Confined to archaea, bacteria and some eukaryotic viruses

[1]The loosely-defined PDX_{9-18} (E/D)XK motif may be present (see text).

NOMENCLATURE CONVENTIONS

Nomenclature of the homing endonucleases is patterned after that of restriction enzymes (15). A three-letter genus-species designation consisting of the first letter of the genus and the first two letters of the species is followed by a Roman numeral to distinguish multiple enzymes from a single organism. Whereas intron endonucleases are characterized by the prefix I- (for intron), the intein endonucleases are characterized by the prefix PI- (for protein insert). Thus, the first-to-be-discovered intron and intein endonucleases of *Saccharomyces cerevisiae* are designated I-*Sce*I and PI-*Sce*I, corresponding to the mitochondrial large rRNA intron ω endonuclease and the nuclear vacuolar-ATPase intein endonuclease, respectively (16,17).

Although the parallel nomenclature conventions for the intron and the intein endonucleases have been useful (16,17), the barrage of recent discoveries of related proteins has raised questions. First, what system should be adopted for those endonucleases that conform to consensus sequence motifs, have demonstrated cleavage activity, but do not reside in introns or inteins? Second, should conventions be developed for putative intron or intein endonucleases that conform to consensus but have not passed functional tests?

In addition to the prefixes I- and PI-, we propose a new prefix, F-, for freestanding, where the endonuclease is not intron- or intein-encoded. For example, according to this convention, the freestanding Endo.*Sce*I and HO endonucleases of *Saccharomyces cerevisiae*, which are members of the LAGLIDADG family, become F-*Sce*I and F-*Sce*II, respectively (Table 2). Likewise, the intergenic SegA and SegE endonucleases of the T-even phage T4, which are members of the GIY-YIG family, become F-*Tev*I and F-*Tev*II, respectively (Table 2). The systematic nomenclature does not preclude maintaining historic names, such as HO and SegA endonuclease, which will continue to be acceptable synonyms.

Table 2. Compilation of homing endonucleases

Name[a]	Alias	Organism[b]	K[c]	Location[d]	Gene/Protein[e]	Intron[f]	Family[g]	Reference
I-*Ani*I	-	*A. nidulans*	E	Mito	cob	I	LAGLIDADG (2)	(77)
I-*Ceu*I	-	*C. eugametos*	E	Chloro	large rRNA	I	LAGLIDADG (1)	(78-80)
I-*Chu*I	-	*C. humicola*	E	Chloro	large rRNA	I	LAGLIDADG (2)	(81)
I-*Cpa*I	-	*C. pallidostigmatica*	E	Chloro	large rRNA	I	LAGLIDADG (1)	(82)
I-*Cpa*II	-	*C. pallidostigmatica*	E	Chloro	small rRNA	I	LAGLIDADG (2)	(41)
I-*Cre*I	-	*C. reinhardtii*	E	Chloro	large rRNA	I	LAGLIDADG (1)	(83)
I-*Csm*I	-	*C. smithii*	E	Mito	cob	I	LAGLIDADG (2)	(84,85)
I-*Dir*I	-	*Di. iridis*	E	Nuclear	small rRNA	I	His-Cys Box	(86)
I-*Dmo*I	-	*De. mobilis*	A	Chromo	large rRNA	A	LAGLIDADG (2)	(87)
I-*Hmu*I	-	*B. subtilis* phage SPO1	B	Phage	DNA pol	I	H-N-H	(38,88)
I-*Hmu*II	-	*B. subtilis* phage SP82	B	Phage	DNA pol	I	H-N-H	(38,89)
I-*Lla*I[h]	-	*L. lactis*	B	Chromo	LtrB	II	H-N-H	(90)
I-*Naa*I	-	*N. andersoni* sp. andersoni	E	Nuclear	small rRNA	I	His-Cys Box	(91)
I-*Por*I	-	*Pyb. organotrophum*	A	Chromo	large rRNA	A	LAGLIDADG (2)	(92)
I-*Ppo*I	-	*Ph. polycephalum*	E	Nuclear[k]	large rRNA	I	His-Cys Box	(93,94)
I-*Sca*I	-	*S. capensis*	E	Mito	cob	I	LAGLIDADG (2)	(18,95)
I-*Sce*I	ω endo	*S. cerevisiae*	E	Mito	large rRNA	I	LAGLIDADG (2)	(96,97)
I-*Sce*II	aI4 endo	*S. cerevisiae*	E	Mito	coxI	I	LAGLIDADG (2)	(98,99)
I-*Sce*III	aI3 endo	*S. cerevisiae*	E	Mito	coxI	I	LAGLIDADG (2)	(100)
I-*Sce*IV	aI5 endo	*S. cerevisiae*	E	Mito	coxI	I	LAGLIDADG (2)	(101,102)
I-*Sce*V[h]	aI2 endo	*S. cerevisiae*	E	Mito	coxI	II	H-N-H	(69)
I-*Sce*VI[h]	aI1 endo	*S. cerevisiae*	E	Mito	coxI	II	H-N-H	(70)
I-*Sce*VII[i]	-	*S. cerevisiae*	E	Mito	cob	I	LAGLIDADG (2)	(18)
I-*Tev*I	-	*E. coli* phage T4	B	Phage	td	I	GIY-YIG	(103,104)
I-*Tev*II	-	*E. coli* phage T4	B	Phage	sunY/nrdD	I	GIY-YIG[l]	(104)
I-*Tev*III	-	*E. coli* phage RB3	B	Phage	nrdB	I	H-N-H	(37)
PI-*Psp*I	-	*Pyc.* sp. GB-D	A	Chromo	DNA pol	-	LAGLIDADG (2)	(105)
PI-*Sce*I	-	*S. cerevisiae*	A	Chromo	VMA1 = TFP1	-	LAGLIDADG (2)	(106,107)
PI-*Tli*I	-	*T. litoralis*	A	Chromo	DNA pol	-	LAGLIDADG (2)	(108)
PI-*Tli*II	-	*T. litoralis*	A	Chromo	DNA pol	-	LAGLIDADG (2)	(20,108)
F-*Sce*[j]	Endo.*Sce*I	*S. cerevisiae*	E	Mito	-	-	LAGLIDADG (2)	(109,110)
F-*Sce*II	HO endo	*S. cerevisiae*	E	Nuclear	-	-	LAGLIDADG (2)	(111)
F-*Suv*I	-	*S. uvarum*	E	Mito	-	-	LAGLIDADG (2)	(11)
F-*Tev*I	segA	*E. coli* phage T4	B	Phage	-	-	GIY-YIG	(112)
F-*Tev*II	segE	*E. coli* phage T4	B	Phage	-	-	GIY-YIG	(113)

[a]Systematic name according to nomenclature conventions.
[b]*A, Asperigillus*; *B, Bacillus*; *C, Chlamydomonas*; *Di, Didymium*; *De, Desulfurococcus*; *E, Escherichia*; *L, Lactococcus*; *N, Naegleria*; *Ph, Physarum*; *Pyb, Pyrobaculum*; *Pyc, Pyrococcus*; *S, Saccharomyces*; *T, Thermococcus*.
[c]K, Kingdom: A, archaea; B, bacteria; E, eukarya.
[d]Chloro, Chloroplast; Chromo, chromosomal; Mito, Mitochondrial.
[e]cob, cytochrome b; cox, cytochrome oxidase; LtrB, relaxase; nrdB, ribonucleotide reductase subunit B; pol, polymerase; rRNA, ribosomal RNA; sunY/nrdD, anaerobic ribonucleotide reductase; td, thymidylate synthase; VMA1, vacuolar membrane ATPase subunit.
[f]Intron type: I, group I; II, group II; A, archaeal.
[g](1), single LAGLIDADG motif; (2), double LAGLIDADG motif.
[h]Associated with RNA subunit.
[i]Endonuclease activity activated by mutation.
[j]Heterodimer with nuclear protein subunit.
[k]Extrachromosomal
[l]The GIY-YIG motif is difficult to discern, as only the YIG is present (at residues 29–31); however BLAST database searches identify extensive similarities downstream of the conserved YIG with the *Neurospora crassa* mitochondrial cob intron 1 and the *Podospora anserina* mitochondrial cob intron 2 GIY-YIG ORFs (S.Pietrokovski, Hutchinson Cancer Research Center, personal communication).

After extensive deliberation with investigators working with homing endonucleases, it has been decided that the nomenclature conventions will continue to be reserved strictly for proteins with demonstrated endonuclease activity. Similarity to one of the four homing endonuclease consensus motifs and/or location within an intron or protein coding sequence will not suffice for an endonuclease designation. Interesting cases for consideration are the yeast maturases, which are intron-encoded LAGLIDADG proteins. Although these proteins are thought to be degenerate endonucleases, they often have no demonstrable DNA cleavage activity, and therefore would not qualify for an endonuclease assignment. However, there is a recent example of nuclease activation by mutation of the maturase. The maturase encoded by the second intron of the cytochrome *b* gene of *S.cerevisiae* can be converted into an active endonuclease by a two amino acid substitution (18). In such circumstances, the wild-type protein is not assigned a formal name, while the mutant variant with demonstrated activity receives an endonuclease designation, namely, I-*Sce*VII (Table 2).

As a result of microbial genome sequencing projects, there is a rapidly growing list of putative inteins. These have recently been compiled and given useful species designations (19). Once

Table 3. Recognition sequences with endonuclease cleavage and intron insertion sites

Name	Recognition sequence[a] ∇		GenBank[b]	Reference[c]
I-*Ani*I	TTATTTGAGGAGG_TTT	C^TCTGTAAATAATGCA	J01387	(77)[d]
I-*Ceu*I	CCGTAACTATAACGGT	C_CTAA^GGTAGCGAAAT	Z17234	(79,80)
I-*Chu*I	GGAAGGTTTGGCAC_CT	CG^ATGTCGGCTCATCG	X68921,X68922	(81,114)
I-*Cpa*I	ATAACGATCCTAAGGT	AG_CGAA^ATTCATTGTC	X68899-X68901	(82,114)}
I-*Cpa*II	GGAATAAGCCCCGGCT	A_ACTC^TGTGCCAGCAG	L39865	(41)
I-*Cre*I	GCTGGGTTCAAAACGT	C_GTGA^GACAGTTTGGT	X01977	(115-117)
I-*Csm*I	TGTACTACCATGGGGT	CAAATGTCTTTCTGGG	X55305	(84)
I-*Dmo*I	AATGCCTTGCCGG_GTA	A^GTTCCGGCGCGCATG	X05480	(87,118)
I-*Hmu*I	GAGTAGTAATGAGCCT	AACG_CTCAGCAATTCC	M37686	(38,88)
I-*Hmu*II	GAGTAGTAATGAGCCT	AACGCTCAACAA(38 nt)A_A	U04812	(38)
I-*Lla*I	TGAACACATCCATAAC^	CATATCATT_TTTAATT	X89922	(90,119)[d]
I-*Por*I	TCCCCGCGAGCCCGTA	A_GGGT^GTGTACGGGGG	M86622	(92,120)
I-*Ppo*I	GTAACTATGACTCTC_T	TAA^GGTAGCCAAATGC	V01159	(93,94)
I-*Sca*I	ATTGTCACATTGAGGT	GCACTAGTTATTACTA	X95974	(18)
I-*Sce*I	AAGTTACGCTAGGG_AT	AA^CAGGGTAATATAGC	V00684	(96,121)
I-*Sce*II	ATTTTGATTCTTTGGT	C_ACCC^TGAAGTATATA	V00694	(98,122)
I-*Sce*III	AATTGGAGGTTTTGG_T	AAC^TATTTATTACCAT	V00694	(98,123)
I-*Sce*IV	ATCTTTTCTCTTG_ATT	A^GCCCTAATCTACGGT	V00694	(98,101)
I-*Sce*V	GTATTAATAATTTTCT^	TCTTAGTAAT_GCCTGC	V00694	(69,98)
I-*Sce*VI	TCACAGTTATTTAATG^	TTTTAGTAGT_TGGTCA	V00694	(70,98)
I-*Sce*VII	ATTGTCACATTGAGGT	GCACTAGTTATTACTA	X95874	(18)
I-*Tev*I	CA_AC^GCTCAGTAGATGTTTTCTTGGGT	CTACCGTTTAATATTG	M12742	(103,104,124)
I-*Tev*II	TTCCAAGCTTATGAGT	ATGAAGTGAACAC_GT^TATTC	Y00122	(104,125)
I-*Tev*III	T^TA_TGTATCTTTTGCGT	GTACCTTTAACTTCCA	X59078	(37)
PI-*Psp*I	CAAAATCCTGGCAAAC	AGCTATTATGGGTATT	U00707	(105)[d]
PI-*Sce*I	TATCTATGTCGG_GTGC^	GGAGAAAGAGGTAATG	J05409	(126)
PI-*Tli*I	GGTTCTTTATGCGG_AC	AC^TGACGGCTTTTATG	M74198	(108)[d]
PI-*Tli*II	TAAATTGCTTGCAAAC	AGCTATTACGGCTATA	M74198	(108)[d]
F-*Sce*I	ACCCTGGATGCTGT_AGGC^ATAGGCTTGGTTAT		J01749	(110)
F-*Sce*II	TTTCAGCTTTCCGC_AACA^GTAAAATTTTATAA		V01313	(111)
F-*Tev*I[e]	ATACGAAACACAAG_A_^A^ATGTTTAGTAAAAC		X03099	(127)
F-*Tev*II	ATTTAATCCTCGCT_TC^AGATATGGCAACTG		X14869	(113,128)

[a]The recognition sequence is loosely defined as the sequence that normally flanks the intron or intein encoding the endonuclease. Sixteen residues to each side of the site are presented, except for those enzymes with distant cut sites. In the case of the freestanding enzymes F-*Sce*I, F-*Sce*II, F-*Tev*I and F-*Tev*II, 14 residues are presented flanking a cleavage site. The length and sequence degeneracy for many of these sites remains to be determined. ∇ and space designate intron or intein insertion site. ^ indicates cleavage on the strand shown; _ indicates cleavage on the complementary strand. Where cleavages are not indicated, they could not be found in the literature. Note that I-*Hmu*I and I-*Hmu*II only cleave one strand.
[b]The GenBank numbers refer to the recognition sequence.
[c]References are to the determination of the cleavage site(s) and the insertion site.
[d]Cleavage-site reference for I-*Ani*I is R. Waring (Temple University), for I-*Lla*I is A. Lambowitz (Ohio State University), for PI-*Psp*I is T. Davis (New England Biolabs), and for PI-*Tli*I and PI-*Tli*II is R. Morgan (New England Biolabs).
[e]F-*Tev*I is reported to cleave at either of two adjacent sites on both strands as indicated.

endonuclease function has been demonstrated, these will simply be assigned PI-prefixes and the appropriate suffix. They will then be included in endonuclease compilations and in REBASE releases (see below), along with restriction enzymes and other homing endonucleases. Nomenclature of other putative homing endonucleases should be treated in similar fashion.

REBASE is a compilation of information about restriction enzymes and methyltransferases that has been maintained for many years by Dr Richard Roberts. REBASE is described in the annual database issue of *Nucleic Acids Research* (7). In addition to providing a literature survey, a great deal of unpublished information is contained within REBASE. Investigators continue to depend upon it to ensure that enzymes are named correctly and to obtain the latest information about which enzymes are available and which DNA sequences they recognize. Recently, homing endonucleases have been included within REBASE. Information from REBASE can be accessed through the World Wide Web (http://www.neb.com/rebase) and can also be received as electronic updates on a monthly schedule.

Dr Roberts (New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA; Tel: +1 508 927 3382; Fax: +1 508 921 1527; Email: roberts@neb.com) will serve as registrar for new homing endonucleases. Email communication is encouraged, and any omissions or errors in the REBASE listing should be registered. A centralized registry for all of the homing endonucleases will ensure not only their inclusion in subsequent REBASE releases, but also will allow systematic assignment of numbers, avoiding confusion in naming successive endonucleases from the same organism. We hope that these nomenclature conventions will help to provide uniformity in the field, particularly as new enzymes are discovered and characterized. Investigators working with and referring to these homing endonucleases are encouraged to use the nomenclature proposed above and to register new enzymes as they are discovered.

RECENT STRUCTURE–FUNCTION INSIGHTS INTO HOMING ENDONUCLEASES

Dissection of protein structure and DNA–protein interaction is proceeding apace for the four families of homing endonucleases. For classification and common sequence features of the four enzyme families the reader is referred to ref. 1 and citations therein. LAGLIDADG endonucleases under intensive study include intron-derived enzymes I-*Cre*I, I-*Cpa*II, I-*Dmo*I, and I-*Por*I, and the intein endonuclease PI-*Sce*I. Also the subject of active investigation are intron endonucleases I-*Tev*I, I-*Ppo*I, and I-*Sce*V, members of the GIY-YIG, His-Cys and H-N-H endonuclease families, respectively. The origin and characteristics of these endonucleases are summarized in Tables 2 and 3, and reviewed in refs. 1–3,8,9,19. Following is a brief review focusing on the literature of the past year. Although there are some clear differences in the architecture and action between the best-studied LAGLIDADG and GIY-YIG families of endonucleases, presumably reflecting their independent ancestry, some common themes also emerge in the properties of these distinctive endonucleases.

Endonuclease and homing-site structure and properties

LAGLIDADG endonucleases. LAGLIDADG enzymes exist with either one or two LAGLIDADG motifs (Table 2) (1), which have been implicated in endonuclease function (20–22). In a major breakthrough for the field, the structures of a single-motif enzyme, I-*Cre*I, and a two-motif enzyme, PI-*Sce*I, have recently been determined (23,24). I-*Cre*I, solved at 3.0 Å resolution, forms a homodimer, with dimensions consistent with its ability to recognize its lengthy homing site, variously estimated at 19–24 bp (24). The 163 residue I-*Cre*I monomers form an elongated protein with a half-cylindrical groove of ~25 × 25 × 35 Å (Fig. 2). The homodimer forms an extended saddle of ~70 Å for DNA binding, with its undersurface consisting of four antiparallel β-strands, which are likely to contact the substrate. The LAGLIDADG motifs are proposed to form the dimer interface, while simultaneously positioning conserved aspartate residues (Asp20 of each monomer) adjacent to the scissile phosphates (Fig. 2). These residues may function to coordinate Mg^{2+}, while conserved arginines are also implicated in catalysis. Considering the symmetry of the dimer, one aspartate from each monomer could allow simultaneous attack across the minor groove to generate 4-nt 3′ overhangs (23).

PI-*Sce*I is a 454-amino acid bifunctional protein, with both endonuclease and protein-splicing activities. Accordingly, the 2.4 Å crystal structure comprises two domains of equivalent size, proposed to correspond to these two functions (24). Domain I, the protein splicing domain, consists almost exclusively of β-sheets, whereas domain II, the endonuclease domain, consists of α/β motifs related by pseudo two-fold symmetry. Statistical modeling of all known inteins has led to the prediction of similar two-domain structures (25). Interestingly, the two α-helices containing the LAGLIDADG motifs in the endonuclease domain of PI-*Sce*I form the axis of symmetry, just as for the I-*Cre*I dimer. The symmetry-related sheets again form the DNA-binding surface, although they generate a platform structure for PI-*Sce*I, rather than a saddle. While LAGLIDADG aspartate residues are also implicated in Mg^{2+} coordination, and a lysine residue is proposed to form part of the active site, the authors argue that a single catalytic center of PI-*Sce*I performs sequential cleavage (24).

From circular permutation analyses it appears that I-*Cre*I binds its homing site substrate largely in the undistorted B-DNA form (24, K.M. Stephens and R. Monat, unpublished). In contrast, PI-*Sce*I bends its lengthy substrate of ~32 bp, as it binds via the major groove and phosphate backbone of the DNA (26,27). Whether differences between I-*Cre*I and PI-*Sce*I are related to their functioning as dimers or monomers, or to whether the endonuclease exists in freestanding form or fused to intein sequences, remains to be determined. In any case, mutational analysis and cocrystal structures will help clarify more exactly the role of the LAGLIDADG motif in catalysis and the precise conformation of the DNA substrates.

Catalysis by PI-*Sce*I is enhanced when a synthetic homing-site duplex is embedded in plasmid DNA, particularly when the plasmid is supercoiled, indicating that the structural and torsional environment of the homing site is important (27). The PI-*Sce*I homing site can be divided into two regions (I and II), both of which are required for cleavage, with the downstream region being a high-affinity binding site necessary for specific complex formation (26). Bipartite recognition sequences may be a common feature of homing endonucleases, as described previously for the yeast mitochondrial LAGLIDADG endonuclease I-*Sce*I (28) and the phage T4 *td*-intron GIY-YIG endonuclease I-*Tev*I (29).

A GIY-YIG endonuclease. The prototypic GIY-YIG endonuclease, I-*Tev*I, binds its substrate as a monomer (30). The homing site is even longer than those of the LAGLIDADG enzymes, approaching 40 bp (31). In contrast to LAGLIDADG endonucleases, where the cleavage and intron insertion site are coincident or separated by a few nucleotides (Table 3), the *td* intron insertion site and I-*Tev*I cleavage sites are separated by 23 and 25 nt (Fig. 3).

On the basis of genetic and biochemical analyses the *td* intron homing site was divided into two regions, with both required for specific cleavage by I-*Tev*I, and with the primary endonuclease recognition domain downstream of the cleavage site, as for PI-*Sce*I (Fig. 3). Tolerance of I-*Tev*I to insertions and deletions between the two domains of the homing site suggested a 'flexible-hinge' model in which these two DNA substrate domains are contacted by two tethered domains of the enzyme. The monomeric enzyme has been separated into two domains, each comprising roughly one half of the 28 kDa enzyme, joined by a protease-sensitive linker (32). The C-terminal domain is the DNA-binding domain, which contacts the primary recognition region of the homing site flanking the intron insertion site. The N-terminal domain is the catalytic domain, which contains the GIY-YIG motif, shown to be important in catalysis, and makes contacts near the cleavage site of the DNA (32). Like the LAGLIDADG enzyme PI-*Sce*I, I-*Tev*I distorts the homing site in the act of binding and catalysis, although for I-*Tev*I interactions are mainly via the minor groove and phosphate backbone. It has also been suggested for both I-*Tev*I (30) and PI-*Sce*I (27) that the naked substrate has some innate structure, which may provide recognition features for the binding of the homing enzymes.

A His-Cys box endonuclease. The His-Cys box contains within a 30-amino acid stretch two conserved histidines and three conserved cysteines, which are likely to form a metal coordination site. I-*Ppo*I is less well characterized than the enzymes described above, but is the best studied of the His-Cys box family of endonucleases. It is a relatively small enzyme (18–20 kDa, depending on translational start site), which acts as a globular dimer and also induces a structural perturbation in the DNA

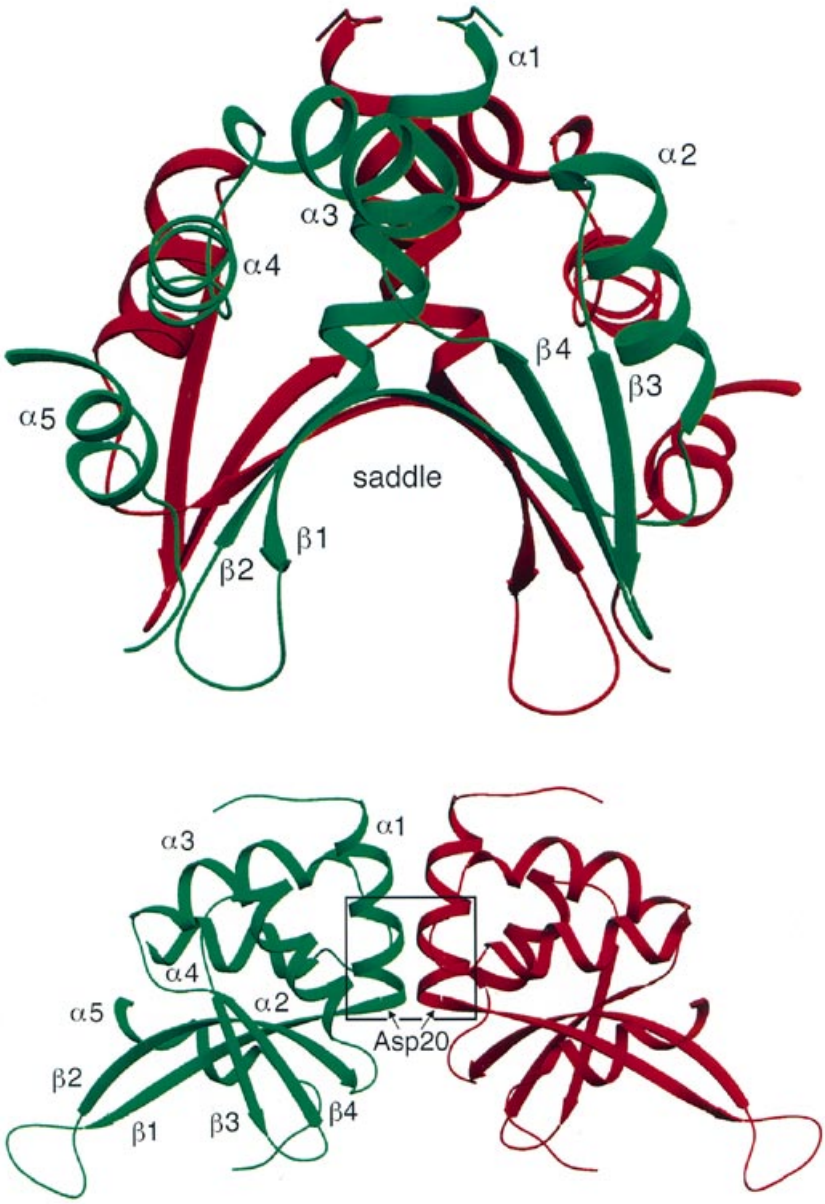

Figure 2. LAGLIDADG endonuclease I-*Cre*I. Two views of the polypeptide backbone are shown, with one monomer colored red, the other green. The top view shows the homodimers viewed down the saddle groove, which is proposed to accommodate the DNA. The bottom view, perpendicular to the saddle groove, shows the LAGLIDADG motif at the α1–β1 boundary, located at the dimer interface, and the asparate residues (ASP20) of each LAGLIDADG motif proposed to be involved in catalysis.

(33,34). Interactions are predominantly through the major groove and appear to be different in two partially symmetrical halves of the recognition sequence (33).

H-N-H endonucleases. The H-N-H proteins contain a consensus sequence spanning 30–33 amino acids, with two pairs of conserved histidines flanking a conserved asparagine to form a zinc finger-like domain (35,36). Although little is known of the structure of the H-N-H endonucleases, they have distinctive and disparate functional properties. While the T-even phage H-N-H enzyme I-*Tev*III makes a double-strand cut, it is the only known homing endonuclease to generate 5′ rather than 3′ extensions (37) (Table 3). Similarly atypical are the *Bacillus subtilis* phage enzymes I-*Hmu*I and I-*Hmu*II, which cleave only one strand of the DNA on both intron-containing and intronless targets (38). More extraordinary yet are the yeast mitochondrial group II intron endonucleases, I-*Sce*V and I-*Sce*VI, which form ribonucleoprotein complexes with their respective intron RNAs. Both the protein and RNA components contribute to cleavage of the DNA substrate. When components of the H-N-H motif of I-*Sce*V were mutated, protein-mediated cleavage of the one strand was blocked, whereas RNA-mediated cleavage of the other strand proceeded normally (Table 3) (12). The colicins and McrA protein from *E.coli* also have an H-N-H motif (1,35,36), but their endonuclease activity has not been well characterized. Nevertheless, the occurrence of the H-N-H motif in these intron-encoded and freestanding nucleases of differing function likely reflects divergent adaptations of a common ancestral H-N-H protein.

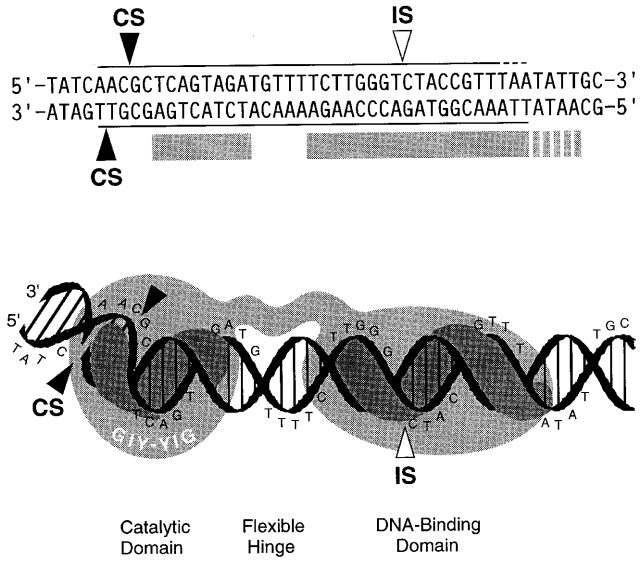

Figure 3. GIY-YIG endonuclease I-*Tev*I. The bipartite homing site is shown above with the two regions of interaction represented by shaded bars, and the minimal homing site is indicated by thin lines. The two-domain enzyme with its DNA-binding and catalytic (GIY-YIG) components joined by a flexible tether is shown below. The endonuclease induces a 40° bend in the cleavage site region of the homing site (29–32). Cleavage site (CS, closed arrowheads); intron insertion site (IS, open arrowhead).

Dynamic interactions and catalysis

The degree to which the details of enzymatic catalysis will be similar for the different homing endonuclease families is unknown. However, interesting themes are apparent in the dynamic properties of the enzyme–DNA complex. These include distortions of the homing site upon endonuclease binding and catalysis, distance sensing as well as site specificity of the enzymes, and product retention by some of the endonucleases.

In general, the distortions induced by the homing endonucleases are in the 40–90° range. Directed bending can promote catalysis in at least three different ways. First, a distortion can facilitate contact between the relatively small proteins and two separated regions of interaction on the DNA as a prerequisite to the transition state, as has been proposed for PI-*Sce*I and I-*Tev*II (27,39). Second, distortion of the substrate can position the scissile phosphates in the active site, as is the case for the restriction enzyme *Eco*RV (40), and as also suggested for PI-*Sce*I (26). Third, a bend toward the major groove can widen the minor groove and facilitate catalysis by allowing access of minor-groove binding enzymes to the scissile phosphates, as has been proposed for I-*Tev*I (30).

Two distortions have been observed for PI-*Sce*I, as well as for I-*Tev*I and I-*Tev*II, although for I-*Tev*I one of these distortions is subtle (26,27,30,39). Interestingly, while DNA is structurally intact in each of the PI-*Sce*I bent complexes, one of the distortions induced by both I- *Tev*I and I-*Tev*II is associated with a nick in one strand. For I-*Tev*I the nick-associated distortion is a directed bend close to the cleavage site. Nicking, which occurs in the absence of added Mg^{2+} for both I-*Tev*I and I-*Tev*II, implies a sequential cleavage mechanism, although cleavage of both strands appears to occur concomitantly in the presence of Mg^{2+}. Preferential nicking of one strand is a property shared by some LAGLIDADG enzymes, as for example I-*Sce*I (28), I-*Chu*I, I-*Ceu*I and I-*Cpa*II (41). Others exhibit concerted cleavage, as for example PI-*Sce*I (27). Enzymes I-*Tev*I, I-*Tev*II, and I-*Cpa*II may resemble *Eco*RV, for which the preference for sequential or concerted cleavage has been linked to divalent cation availability (30,41,42). It is unclear for the homing enzymes whether sequential versus concerted cleavage represents a mechanistic or simply a kinetic difference. The ways in which divalent cations and DNA distortions promote catalysis in these different enzyme systems must await further biochemical analysis and structure determination of the enzyme–substrate complexes.

Another feature of the GIY-YIG protein I-*Tev*I that may be common to the LAGLIDADG enzymes is manifest once binding has been established. Genetic experiments indicate that I-*Tev*I selects its cut site by both distance sensing and sequence discrimination (31,43). Based on a comparison of homing sites, a similar cleavage-site selection mechanism has been proposed for PI-*Sce*I (26). Furthermore, both enzymes, like I-*Sce*I, F-*Sce*II and I-*Tev*II, remain bound to one cleavage product (27,39,44,45). In contrast, both I-*Por*I and I-*Ppo*I appear to be released after catalysis (cited in 27). Persistent binding of the endonuclease to one of the cleavage products can have genetic consequences in the ensuing recombination events (44).

It is noteworthy that I-*Tev*I, I-*Tev*II, F-*Tev*I, and F-*Tev*II, all GIY-YIG endonucleases, generate 2-nt 3′ extensions, whereas all LAGLIDADG enzymes characterized to date leave 4-nt 3′ extensions. A classification of restriction endonuclease structures on the basis of cleavage pattern (i.e., the nature and length of single-strand extensions) has been proposed (4,46,47). Given the foregoing, it will be of interest to see whether for the GIY-YIG, LAGLIDADG, and other families of nucleases, the position of scissile phosphates is also a correlative feature with the structure and catalytic properties of the homing enzymes.

UTILITY OF HOMING ENDONUCLEASES IN GENE MANIPULATION AND RECOMBINATION RESEARCH

Rare-cutting endonucleases allow one to introduce one or a few double-strand breaks into complex genomes. This capability makes the homing enzymes useful tools for analyzing and manipulating genomes for mapping, gene cloning and targeting, and for studying double-strand-break (DSB) repair in diverse biological systems.

Genome analysis and gene manipulation

Mapping. Genome mapping strategies have been based both on naturally available cleavage sites and on the introduction of cleavage sites as chromosomal landmarks. The homing endonucleases have been used in combination with rare-cutting restriction enzymes to map a variety of bacterial genomes, and have been particularly useful for analyzing chromosomal organization. I-*Ceu*I, for example, cleaves only in the rRNA genes of many bacterial strains, which harbor multiple copies of the gene. Mapping I-*Ceu*I fragments therefore allows one to probe genome configuration in the rDNA region. Such approaches have underscored chromosomal plasticity in several bacterial species (48,49).

Mapping has also been achieved by introduction of novel sites into genomes, with transposons engineered to contain cleavage sites. These approaches have been used for the study of genome organization, and for chromosome fragmentation for cloning large DNA fragments in bacteria and yeast. Available systems

include mini-Tn10::I-*Sce*I and Tn5::I-*Sce*I cassettes for use in Gram-negative bacteria (50–52) and a retrotransposon-based Ty1::I-*Dmo*I cartridge for use in yeast (53). The Ty1 system has been useful in analyzing both native yeast chromosomes, and mouse genes in yeast artificial chromosomes (YACs). Genetically engineered I-*Sce*I sites in yeast have also been helpful for physical mapping of yeast contigs for yeast genome sequencing (54).

Cloning. Vectors have been developed for cloning large fragments generated by homing endonucleases. These include plasmid vectors with a multiple cloning site containing homing-endonuclease cleavage sites for PI-*Sce*I, I-*Ppo*I, I-*Ceu*I, and PI-*Tli*I (55). Cosmid vectors with an I-*Sce*I site are also available (56).

Targeting. Double-strand breaks (DSBs) are recombinogenic, facilitating both homologous and non-homologous recombination events. It has been shown that homologous recombination can be stimulated 10- to 1000-fold in both pro- and eukaryotic systems by a DSB. This phenomenon provides a means for targeting integration events from a transformed or transfected sequence to the chromosome by introduction of a DSB in a homologous sequence on the genome. Such gene-targeting strategies have been facilitated by the expression of homing endonucleases in fungal, plant, and mammalian cells (reviewed in 57) or by electroporation of purified enzyme into cells (58). The existence or introduction of sites at defined positions within genomes therefore creates the ability to engineer targeted deletions or insertions into genomes in many different biological systems. Indeed, gene targeting has been achieved in embryonic stem cells expressing I-*Sce*I, paving the way to create transgenic animals by homing endonuclease-directed gene targeting (reviewed in 57).

Homing endonucleases in studies of DSB repair

DSBs must be repaired in all organisms to maintain chromosomal integrity and viability. DSB repair also plays a role in DNA rearrangements such as intron mobility, in both prokaryotic and eukaryotic systems (reviewed in 1,9). Furthermore, mating-type switching in fungi (59), transposition in flies (60), and V(D)J recombination in mammalian cells (61) are all DSB-dependent events. The highly specific homing endonucleases provide the ability to direct discrete breaks into genomes generating isolated foci for study of both homologous and non-homologous DSB-repair events.

Homology-dependent events. Homing endonucleases have been used to study homology-dependent DSB-repair pathways in phage T4 (62,63), yeast (2,64,65), plants (66,67), and mammalian cells (57,68). Not only have these studies shed light on the functional requirements of the repair events, but they have illuminated different recombination pathways. One emerging theme from these studies has been the tight coupling of DNA replication and DSB repair in gene conversion events in which foreign sequences are used to repair the breaks, as in group I intron homing. These studies have taken advantage of I-*Tev*I in the phage system and the HO endonuclease, F-*Sce*II, in yeast (62–64).

Another area in which we have gained new insight is in RNA-dependent group II intron homing, also called retrohoming, with I-*Sce*V and I-*Sce*VI. As part of RNP complexes, these remarkable intron-encoded proteins, which have reverse transcriptase and RNA maturase function in addition to endonuclease activity, are physically associated with the excised intron RNA. The RNA cleaves the sense strand of the DNA homing site by reverse splicing while the protein cleaves the antisense strand, to generate a primer for reverse transcription of the intron RNA (12,69). Alternatively, the intron RNA can insert itself directly into double-stranded DNA (70). In either event, repair occurs via a cDNA copy of the intron RNA.

Homology-independent events. Homing endonucleases have also been used to study homology-independent events in bacterial/phage (71), fungal (72,73), and mammalian systems (57,68). An interesting finding emerged from studies with the HO-endonuclease, F-*Sce*II, to initiate DSB repair in *S.cerevisiae* in which homologous recombination had been inhibited (72,73). Under such conditions most of the DSBs were repaired by end-joining events similar to those found in mammalian cells, whereas ~1% of the events reflected capture of cDNAs corresponding to Ty1 retrotransposon mRNA. While such events have some features in common with group II retrohoming, they provide a possible mechanism for insertion of pseudogenes and short and long interspersed nuclear sequences (SINEs and LINEs) in eukaryotic genomes. Clearly, the repair of DSBs by foreign DNAs, including endogenous retroelements, is important in the evolution of genomes.

EVOLUTION OF HOMING ENDONUCLEASES

In considering the evolution of homing endonucleases one must address questions at both the structure-function level, and at the level of the persistence of these apparently discretionary elements in biological systems.

Evolution of endonuclease structure

It has been proposed that the double LAGLIDADG motif homing endonucleases evolved from the single-motif enzymes by a gene duplication event. In support of this argument are protein footprinting experiments of the two-motif archaeal endonucleases I-*Dmo*I and I-*Por*I. The results suggest that these enzymes consist of two repeats with each containing one LAGLIDADG motif (74). Furthermore, single-motif enzymes like I-*Cre*I act as dimers on pseudopalindromic substrates (23), whereas double-motif enzymes like PI-*Sce*I bind as monomers on asymmetric substrates (27). The gene duplication hypothesis is lent further credence by the structure of these enzymes. The substrate-binding surface of I-*Cre*I is created by the symmetric juxtaposition of the monomers about the LAGLIDADG motifs; in much the same way, the DNA-binding structure in the nuclease domain of PI-*Sce*I has a pseudo 2-fold symmetry about its two LAGLIDADG motifs (23,24). It has been postulated that the derived two-motif duplicated monomers have evolved a relaxed requirement for symmetry, thereby allowing the enzymes to acquire an expanded substrate repertoire (26,74).

Whereas in the above scenario binding and catalytic regions would be interdigitated in each half of the double-motif LAGLIDADG enzymes, they are separated by a flexible tether in the monomeric GIY-YIG enzyme I-*Tev*I (Fig. 3). Although the GIY-YIG motif that forms the catalytic domain of I-*Tev*I is conserved in different GIY-YIG proteins, the DNA-binding region is variant. It has therefore been proposed that the GIY-YIG domain is a catalytic cartridge that can be combined with different

DNA-binding proteins to evolve nucleases with altered specificities (32).

Endonuclease persistence in diverse organisms

Homing endonuclease genes have been considered highly invasive elements that gain access into genomes by virtue of the ability of their products to make DSBs and promote recombination. Their propagation is ensured when these parasitic elements find refuge in introns and inteins (reviewed in 8). While their sheer invasiveness would secure the persistence and dissemination of endonuclease genes in biological systems, there have also been examples of the selective advantage to organisms with intron-encoded endonucleases. These include the ability of the phage SP82 intron endonuclease to exclude genetic markers of related phage in mixed infections (38), and the selective advantage of an archaeal rDNA intron to *Sulfolobus acidocaldarius* (75).

While restriction enzymes may similarly be of advantage to bacteria through their ability to limit phage infection, restriction–modification systems have also been shown to behave in a selfish manner (76). Why then do the intron and intein endonucleases engage in self-propagating homing reactions, whereas restriction enzymes do not? One possibility is related to the rarity of homing endonuclease cut sites. On the one hand, the action of the frequently cutting restriction enzymes is precluded *in vivo* by their cognate modifying enzymes, except with foreign unmodified DNA, which is likely to be degraded by the enzyme and unable to perpetuate a homing event. On the other hand, the recognition site of the homing endonucleases is so large that there is likely to be only one site per genome; once cleaved and occupied by the endonuclease-encoding ORF, further cleavage at this site would be prevented, while homing to similar unoccupied sites would be ensured. A second possibility is related to the tendency of homing endonucleases to tenaciously bind their cleavage products via their lengthy recognition sequences and thereby influence subsequent recombination events (44). Regardless of why homing endonucleases promote DNA rearrangements, their ability to do so is an important factor in the evolution of genomes.

ACKNOWLEDGEMENTS

We are thankful to Barry Stoddard, Fred Gimble and Florante Quiocho for providing endonuclease structures before publication, and Shmuel Pietrokovski and Richard Waring for sharing unpublished data. We are also grateful to Mary Bryk, Elaine Davis, Vicky Derbyshire, Fred Gimble, Debbie Hinton, Claude Jacq, Claude Lemieux, Alan Lambowitz, Ray Monat, Fran Perler, Phil Perlman, Shmuel Pietrokovski, Alfred Pingoud, David Shub, Barry Stoddard, Jeremy Thorner, Monique Turmel and members of the Belfort Laboratory for their constructive suggestions on this review. Thanks also to Maureen Belisle, George Silva and Patrick VanRoey for help with the figures, and to Maryellen Carl for preparing the manuscript. Work in the authors' laboratories is supported by grants from the NIH, GM39422 and GM44844 to MB, and LM04971 to RR.

REFERENCES

1 Belfort, M. and Perlman, P.S. (1995) *J. Biol. Chem.* **270**, 30237–30240.
2 Curcio, M.J. and Belfort, M. (1996) *Cell* **84**, 9–12.
3 Cooper, A.A. and Stevens, T.H. (1995) *Trends Biochem.* **20**, 351–356.
4 Anderson, J.E. (1993) *Curr. Opin. Struct. Biol.* **3**, 24–30.
5 Aggarwal, A.K. (1995) *Curr. Opin. Struct. Biol.* **5**, 11–19.
6 Pingoud, A. and Jeltsch, A. (1997) *Eur. J. Biochem.* **246**, 1–22.
7 Roberts, R.J. and Macelis, D. (1997) *Nucleic Acids Res.* **25**, 248–262.
8 Mueller, J.E., Bryk, M., Loizos, N. and Belfort, M. (1993) In Linn, S.M., Lloyd, R.S. and Roberts, R.J., eds, *Nucleases* 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 111–143.
9 Lambowitz, A.M. and Belfort, M. (1993) *Annu. Rev. Biochem.* **62**, 587–622.
10 Roberts, R.J. and Halford, S.E. (1993) In Linn, S.M., Lloyd, R.S. and Roberts, R.J., eds, *Nucleases* 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 35–88.
11 Shibata, T., Nakagawa, K. and Morishima, N. (1995) *Adv. Biophys.* **31**, 77–91.
12 Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S. and Lambowitz, A. (1995) *Cell* **83**, 529–538.
13 Bickle, T.A. (1993) In Linn, S.M., Lloyd, R.S. and Roberts, R.J., eds, *Nucleases* 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 89–109.
14 Wilson, G.G. (1988) *Gene* **74**, 281–289.
15 Smith, H.O. and Nathans, D.A. (1973) *J. Mol. Biol.* **81**, 419–423.
16 Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S. and Vogt, V.M. (1989) *Gene* **82**, 115–118.
17 Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J. and Belfort, M. (1994) *Nucleic Acids Res.* **22**, 1125–1127.
18 Szczepanek, T. and Lazowska, J. (1996) *EMBO J.* **15**, 3758–3767.
19 Perler, P.B., Olsen, G.J. and Adam, E. (1997) *Nucleic Acids Res.* **25**, 1087–1093.
20 Hodges, R.A., Perler, F.B., Noren, C.J. and Jack, W.E. (1992) *Nucleic Acids Res.* **20**, 6153–6157.
21 Gimble, F.S. and Stephens, B.W. (1995) *J. Biol. Chem.* **270**, 5849–5856.
22 Henke, R.M., Butow, R.A. and Perlman, P.S. (1995) *EMBO J.* **14**, 5094–5099.
23 Health, P.J., Stephens, K.M., Monnat, R.J., and Stoddard, B.L. (1997) *Nature Struct. Biol.* **4**, 468–476.
24 Duan, X., Gimble, F.S., and Quiocho, F.A. (1997) *Cell* **89**, 555–564.
25 Dalgaard, J.Z., Moser, M.J., Hughey, R., and Mian, I.S. (1997) *J. Comput. Biol.* (in press).
26 Gimble, F.S. and Wang, J. (1996) *J. Mol. Biol.* **263**, 163–180.
27 Wende, W., Grindl, W., Christ, F., Pingoud, A. and Pingoud, V. (1996) *Nucleic Acids Res.* **24**, 4123–4132.
28 Perrin, A., Buckle, M. and Dujon, B. (1993) *EMBO J.* **12**, 2939–2947.
29 Bryk, M., Quirk, S.M., Mueller, J.E., Loizos, N., Lawrence, C. and Belfort, M. (1993) *EMBO J.* **12**, 2141–2149.
30 Mueller, J.E., Smith, D., Bryk, M. and Belfort, M. (1995) *EMBO J.* **14**, 5724–5735.
31 Bryk, M., Belisle, M., Mueller, J.E. and Belfort, M. (1995) *J. Mol. Biol.* **247**, 197–210.
32 Derbyshire, V., Kowalski, J.C., Dansereau, J.T., Hauer, C.R. and Belfort, M. (1997) *J. Mol. Biol.* **265**, 494–506.
33 Ellison, E.L. and Vogt, V.M. (1993) *Mol. Cell Biol.* **13**, 7531–7539.
34 Wittmayer, P.K. and Raines, R.T. (1996) *Biochemistry* **35**, 1076–1083.
35 Gorbalenya, A.E. (1994) *Protein Sci.* **3**, 1117–1120.
36 Shub, D.A., Goodrich-Blair, H. and Eddy, S.R. (1994) *Trends Biochem.* **19**, 402–404.
37 Eddy, S.R. and Gold, L. (1991) *Genes Dev.* **5**, 1032–1041.
38 Goodrich-Blair, H. and Shub, D.A. (1996) *Cell* **84**, 211–221.
39 Loizos, N., Silva, G.H. and Belfort, M. (1996) *J. Mol. Biol.* **255**, 412–424.
40 Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) *EMBO J.* **12**, 1781–1795.
41 Turmel, M., Mercier, J.-P., Cote, V., Otis, C. and Lemieux, C. (1995) *Nucleic Acids Res.* **23**, 2519–2525.
42 Halford, S.E. and Goodall, A.J. (1988) *Biochemistry* **27**, 1771–1777.
43 Bell-Pedersen, D., Quirk, S.M., Bryk, M. and Belfort, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7719–7723.
44 Mueller, J.E., Smith, D. and Belfort, M. (1996) *Genes Dev.* **10**, 2158–2166.
45 Jin, Y., Binkowski, G., Simon, L.D. and Norris, D. (1997) *J. Biol. Chem.* **272**, 7352–7359.
46 Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S. and Kokkinidis, M. (1994) *Struct. Biol.* **1**, 469–475.
47 Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994) *Nature* **368**, 660–664.

48 Liu, S.L. and Sanderson, K.E. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10303–10308.
49 Toda, T. and Itaya, M. (1995) *Microbiology* **141**, 1937–1945.
50 Bloch, C.A., Rode, C.K., Obreque, V.H. and Mahillon, J. (1996) *Biochem. Biophys. Res. Commun.* **223**, 104–111.
51 Mahillon, J., Rode, C.K., Leonard, C. and Bloch, C.A. (1997) *Gene* **187**, 273–279.
52 Jumas-Bilak, E., Maugard, C., Michaux-Charachon, S., Allardet-Servent, A., Perrin, A., O'Callaghan, D. and Ramuz, M. (1995) *Microbiology* **141**, 2425–2432.
53 Dalgaard, J.Z., Banerjee, M. and Curcio, M.J. (1996) *Genetics* **143**, 673–683.
54 Thierry, A., Gaillon, L., Galibert, F. and Dujon, B. (1995) *Yeast* **11**, 121–135.
55 Asselbergs, F.A.M. and Rival, S. (1996) *BioTechniques* **20**, 558–562.
56 Favre, D. and Viret, J.F. (1996) *Gene* **178**, 43–49.
57 Jasin, M. (1996) *Trends Genet.* **12**, 224–228.
58 Brenneman, M., Gimble, F.S. and Wilson, J.H. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 3608–3612.
59 Haber, J.E. (1992) *Trends Genet*. **8**, 446–452.
60 Nassif, N., Penney, J., Pal, S., Engels, W.R. and Gloor, G.B. (1994) *Mol. Cell. Biol*. **14**, 1613–1625.
61 Gellert, M. (1994) *Semin. Immunol.* **6**, 125–130.
62 Mueller, J.E., Clyman, J., Huang, Y., Parker, M.M. and Belfort, M. (1996) *Genes Dev.* **10**, 351–364.
63 George, J.W. and Kreuzer, K.N. (1996) *Genetics* **143**, 1507–1520.
64 Malkova, A., Ivanov, E.L. and Haber, J.E. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 7131–7136.
65 Nelson, H.H., Sweetser, D.B. and Nickoloff, J.A. (1996) *Mol. Cell. Biol.* **16**, 2951–2957.
66 Puchta, H., Dujon, B. and Hohn, B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5055–5560.
67 Durrenberger, F., Thompson, A.J., Herrin, D.L. and Rochaix, J.D. (1996) *Nucleic Acids Res*. **24**, 3323–3331.
68 Liang, F., Romanienko, P.J., Weaver, D.T., Jeggo, P.A. and Jasin, M. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8929–8933.
69 Zimmerly, S., Guo, H., Perlman, P.S. and Lambowitz, A.M. (1995) *Cell* **82**, 545–554.
70 Yang, J., Zimmerly, S., Perlman, P.S. and Lambowitz, A.M. (1996) *Nature* **381**, 332–335.
71 Parker, M.M., Court, D.A., Preiter, K. and Belfort, M. (1996) *Genetics* **143**, 1057–1068.
72 Moore, J.K. and Haber, J.E. (1996) *Nature* **383**, 644–646.
73 Teng, S.-C., Kim, B. and Gabriel, A. (1996) *Nature* **383**, 641–644.
74 Lykke-Andersen, J., Garrett, R.A. and Kjems, J. (1996) *Nucleic Acids Res.* **24**, 3982–3989.
75 Aagaard, C., Dalgaard, J.Z. and Garrett, R.A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 12285–12289.
76 Naito, T., Kusano, K. and Kobayashi, I. (1995) *Science* **267**, 897–899.
77 Waring, R.B., Davies, R.W., Scazzocchio, C. and Brown, T.A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6332–6336.
78 Gauthier, A., Turmel, M. and Lemieux, C. (1991) *Curr. Genet.* **19**, 43–47.
79 Marshall, P. and Lemieux, C. (1991) *Gene* **104**, 241–245.
80 Turmel, M., Boulanger, J., Schnare, M.N., Gray, M.W. and Lemieux, C. (1991) *J. Mol. Biol.* **218**, 293–311.
81 Cote, V., Mercier, J.-P., Lemieux, C. and Turmel, M. (1993) *Gene* **129**, 69–76.
82 Turmel, M., Cote, V., Otis, C., Mercier, J.-P., Gray, M.W., Lonergan, K.M. and Lemieux, C. (1995) *Mol. Biol. Evol.* **12**, 533–545.
83 Rochaix, J.D., Rahire, M. and Michel, F. (1985) *Nucleic Acids Res*. **13**, 975–984.
84 Colleaux, L., Michel-Wolwertz, M.-R., Matagne, R.F. and Dujon, B. (1990) *Mol. Gen. Genet*. **223**, 288–296.
85 Ma, D.-P., King, Y.-T., Kim, Y. and Luckett Jr., W.S. (1992) *Plant Mol. Biol.* **18**, 1001–1004.
86 Johansen, S., Embley, T.M. and Willassen, N.P. (1993) *Nucleic Acids Res.* **21**, 4405
87 Dalgaard, J.Z., Garrett, R.A. and Belfort, M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5414–5417.
88 Goodrich-Blair, H., Scarlato, V., Gott, J.M., Xu, M. and Shub, D.A. (1990) *Cell* **63**, 417–424.
89 Goodrich-Blair, H. and Shub, D.A. (1994) *Nucleic Acids Res*. **22**, 3715–3721.
90 Mills, D.A., McKary, L.L. and Dunny, G.M. (1996) *J. Bacteriol*. **178**, 3531–3538.
91 Embley, T.M., Dyal, P. and Kilvington, S. (1992) *Nucleic Acids Res*. **20**, 6411.
92 Lykke-Andersen, J., Thi-Ngoc, H.P. and Garrett, R.A. (1994) *Nucleic Acids Res.* **22**, 4583–4590.
93 Muscarella, D.E., Ellison, E.L., Ruoff, B.M. and Vogt, V.M. (1990) *Mol. Cell Biol*. **10**, 3386–3396.
94 Muscarella, D.E. and Vogt, V.M. (1989) *Cell* **56**, 443–454.
95 Lazowska, J., Szczepanek, T., Macadre, C. and Dokova, M. (1992) *C. R. Acad. Sci. Paris* **315**, 37–41.
96 Dujon, B. (1980) *Cell* **20**, 185–197.
97 Colleaux, L., D'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. and Dujon, B. (1986) *Cell* **44**, 521–533.
98 Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) *J. Biol. Chem*. **255**, 11927–11941.
99 Hanson, D.K., Lamb, M.R., Mahler, H.R. and Perlman, P.S. (1982) *J. Biol. Chem*. **257**, 3218–3224.
100 Sargueil, B., Delahodde, A., Hatat, D., Tian, G.L., Lazowska, J. and Jacq, C. (1991) *Mol. Gen. Genet.* **225**, 340–341.
101 Moran, J.V., Wernette, C.M., Mecklenburg, K.L., Butow, R.A. and Perlman, P.S. (1992) *Nucleic Acids Res*. **20**, 4069–4076.
102 Seraphin, B., Faye, G., Hatat, D. and Jacq, C. (1992) *Gene* **113**, 1–8.
103 Chu, F.K., Maley, G., Pedersen-Lane, J., Wang, A.-M. and Maley, F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3574–3578.
104 Bell-Pedersen, D., Quirk, S., Clyman, J. and Belfort, M. (1990) *Nucleic Acids Res*. **18**, 3763–3770.
105 Xu, M., Southworth, M.W., Mersha, F.B., Hornstra, L.J. and Perler, F.B. (1993) *Cell* **75**, 1371–1377.
106 Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. and Anraku, Y. (1990) *J. Biol. Chem*. **265**, 6726–6733.
107 Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) *Science* **250**, 651–657.
108 Perler, F.B., Comb, D.G., Jack, W.E., Moran, L.S., Quiang, B., Kucera, R.B., Benner, J., Slatko, B.E., Nwankwo, D.O., Hempstead, S.K., Carlow, C.K.S. and Jannasch, H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5577–5581.
109 Watabe, H., Shibata, T. and Ando, T. (1981) *J. Biochem.* **90**, 1623–1632.
110 Watabe, H., Iino, T., Kaneko, T., Shibata, T. and Ando, T. (1983) *J. Biol. Chem.* **258**, 4663–4665.
111 Kostriken, R., Strathern, J.N., Klar, A.J.S., Hicks, J.B. and Heffron, F. (1983) *Cell* **35**, 167–174.
112 Sharma, M., Ellis, R.L. and Hinton, D.M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6658–6662.
113 Kadyrov, F.A., Kryukov, V.M., Shlyapnikov, M.G. and Baev, A.A. (1994) *Doklady Biochem.* **339**, 145–147.
114 Turmel, M., Gutell, R.R., Mercier, J.-P., Otis, C. and Lemieux, C. (1993) *J. Mol. Biol.* **232**, 446–467.
115 Durrenberger, F. and Rochaix, J.-D. (1993) *Mol. Gen. Genet.* **236**, 409–414.
116 Thompson, A.J., Yuan, X., Kudlicki, W. and Herrin, D.L. (1992) *Gene* **119**, 247–251.
117 Allet, B. and Rochaix, J.-D. (1979) *Cell* **18**, 55–60.
118 Kjems, J. and Garrett, R.A. (1985) *Nature* **318**, 675–677.
119 Shearman, C., Godon, J.-J. and Gasson, M. (1996) *Mol. Microbiol.* **21**, 45–53.
120 Dalgaard, J.Z. and Garrett, R.A. (1992) *Gene* **121**, 103–110.
121 Colleaux, L., D'Auriol, L., Galibert, F. and Dujon, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
122 Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) *Cell* **56**, 431–441.
123 Perea, J., Desdouets, C., Schapria, M. and Jacq, C. (1993) *Nucleic Acids Res*. **21**, 358.
124 Bell-Pedersen, D., Quirk, S.M., Aubrey, M. and Belfort, M. (1989) *Gene* **82**, 119–126.
125 Shub, D.A., Gott, J.M., Xu, M.-Q., Lang, B.F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. and Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1151–1155.
126 Gimble, F.S. and Thorner, J. (1992) *Nature* **357**, 301–306.
127 Sharma, M. and Hinton, D.M. (1994) *J. Bacteriol*. **176**, 6439–6448.
128 Kaliman, A.V., Khasanova, M.A., Kryukov, V.M., Tanyashin, V.I. and Bayev, A.A. (1990). *Nucleic Acids Res*. **18**, 4277.