

This collection of paper citations, their abstracts and a few editorial comments summarizes many (but not all!) of the seminal studies of homing endonuclease genetics and functions, from 1978 to 1995, culminating in their first application for targeted genetic modification studies (at sites containing pre-integrated LHE target sites) in 1994 and 1995.

Pre-1980:

Bos et al, An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA.. Nature. 1978 Sep 28;275(5678):336-8.

This paper describes the initial visualization of intervening sequence (split gene structure) of 23S rDNA from yeast mitochondrial genome, using electron microscopy on cDNA hybridized with rRNA probe. The intervening sequence had previously been observed to display dominant, non-Mendelian inheritance in yeast mating experiments (a phenomena that eventually became known and recognized as 'intron homing'). This paper was published about a year after the initial observation of split gene structures, using similar techniques, by Hogness, Roberts and Sharp in separate papers (Roberts and Sharp later won the Nobel prize for the discovery of split genes).

Faye et al, Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast *Saccharomyces cerevisiae*.. Mol Gen Genet. 1979 Jan 5;168 (1):101-9.

The sequence organization of the yeast mit-DNA region carrying the large ribosomal RNA gene and the polar locus omega was examined. Hybridization studies using rho- deletion mutants and electron microscopy of the heteroduplexes formed between 23S rRNA and the appropriate restriction fragments, lead to the conclusion that the 23S rRNA1 gene of the omega+ strains is split by an insertion sequence of 1,000-1,100 bp. In contrast, no detectable insertion was found in the 23S rRNA gene of the omega- strains. The size and the location of the insert found in the 23S rRNA gene of the omega+ strains appear to be identical to those of the sequence delta which had previously been found to characterize the difference (at the omega locus) between the mitDNA of the wild type strains carrying the omega+ or omega- alleles.

Notes: These two papers describe the observation that the genetic loci termed 'omega', which was previously observed in mitochondrial crosses to be dominantly inherited and persistent (in internal bulletins at the Pasteur Institute in 1971 and after), corresponds to an intervening sequence in the yeast mitochondrial 23S rDNA.

1980:

Dujon, et al. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci.. Cell. 20: 185 - 197. (1980)

The complete nucleotide sequence has been determined for the intron, its junctions and the flanking exon regions of the 21S rRNA gene in three genetically characterized strains differing by their omega alleles (omega+, omega- and omega n) and by their chloramphenicol-resistant mutations at the rib-1 locus. Comparison of these DNA sequences shows that: --omega+ differs from omega- and omega n by the presence of the intron (1143 bp), as well as by a second and unexpected mini-insert (66 bp) located 156 bp upstream within the exon, whose nature and functions are still unknown but whose striking palindromic structure may suggest a mitochondrial transposable element. --The two mutations C321R and C323R correspond to two different monosubstitutions, 56 bp apart in the omega- and omega n strains but separated by the intron in the omega+ strains. In relation to previous genetic results, a model is discussed assuming that the interactions of two different regions or genetic loci determine the chloramphenicol resistance, one of which contains the omega n mutations. --A long uninterrupted coding sequence able to specify a 235 amino acid polypeptide exists within the intron. This remarkable observation gives new insight into the origin of the mitochondrial introns and raises the question of the possible functions of intron-encoded polypeptides. Finally, sequence comparisons with evolutionarily distant organisms, showing that different rRNA introns are inserted at different positions of an otherwise highly conserved region of the gene, suggest a recent insertion of these introns and a mechanism for splicing after the assembly of the large ribosomal subunit.

Lazowska et al. Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron.. Cell. 22: 333 - 348. (1980)

We have determined the DNA sequence of the wild type and mutated introns as well as their flanking exons in the yeast mitochondrial gene specifying cytochrome b. The second intron (box3) encodes a trans-acting protein "mRNA maturase" responsible for splicing and maturation of cytochrome b mRNA. This protein is interlaced with cytochrome b exon sequences. Its biosynthesis is subject to a negative feedback which may constitute a regulatory mechanism for the expression of split genes.

Notes: These two papers are the first to describe, at the level of coding sequence and corresponding amino acid sequence, the existence of intron-encoded proteins, and led to the hypotheses that these proteins might be involved in either intron mobility (Dujon) or intron splicing (Lazowska).

1983:

Klostriken et al., A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. Cell. 35: 167 - 174. (1983)

We have detected two site-specific endonucleases in strains of *Saccharomyces cerevisiae*. One endonuclease, which we call YZ endo, is present only in yeast strains that are undergoing mating-type interconversion. The site at which YZ endo cleaves corresponds to the in vivo double-strand break occurring at the mating-type locus in yeast undergoing mating-type interconversion. YZ endo generates a site-specific double-strand break having 4-base 3' extensions terminating in 3' hydroxyl groups. The site of cleavage occurs in the Z1 region near the YZ junction of the mating-type locus. Mutant mating-type loci known to decrease the frequency of mating-type interconversion are correspondingly poor substrates for YZ endo in vitro. In vitro analysis of a number of such altered recognition sites has delimited the sequences required for cleavage. The molecular genetics of mating-type interconversion is discussed in the context of this endonucleolytic activity. The second endonuclease, which we refer to as Sce II, is present in all strains of *S. cerevisiae* we have examined. The cleavage site of Sce II has been determined and proves to be unrelated to the cleavage site of YZ endo.

'YZ endo' corresponds to the HO endonuclease. This is a free-standing endo of the LAGLIDADG family that also contains remnants of a devolved intein domain. It is structurally related to the intein-associated endos such as PI-SceI; both domains are involved in DNA recognition.

This paper describes the discovery of the endonuclease that cuts in the MAT locus (that encodes a transcription factor which activates either of two different 'mating type' suites of genes), thereby stimulating recombination with either of the 'Hidden MAT' loci. This gene conversion event results in conversion of the MAT transcription factor type between 'a' and 'α'. The study includes a description of the many of the hallmarks of LAGLIDADG DNA recognition: the target site sequence, a lower estimate of its length (greater than 12 bp) and the observation of variable fidelity at some positions.

The structure of HO endo was later modeled and used to conduct a mutagenesis study of its function (Bakhrat, A., Jurica, M. S., **Stoddard, B. L.** and Raveh, D. (2004) "Homology modeling and mutational analysis of HO endonuclease of yeast" *Genetics* **166** (2) 721 – 728).

Hensgens et al., Two intron sequences in yeast mitochondrial COX1 gene: Homology among URF-containing introns and strain-dependent variation in flanking exons. Cell. 32: 379-389 (1983)

The DNA sequences of two optional introns in the gene for subunit I of cytochrome c oxidase in yeast mitochondrial DNA have been determined. Both contain long unassigned reading frames (URFs). These display regions of amino acid homology with six other URFs, two of which encode proteins involved in mitochondrial RNA splicing. Such conserved regions may thus define functionally important domains of proteins involved in RNA processing. This homology also implies that these URFs had a common ancestral sequence, which has been duplicated and dispersed around the genome. Comparison of the flanking exons in the long strain KL14-4A with their unsplit counterpart in D273-10B reveals clustered sequence differences, which lead in D273-10B to codons rarely used in exons. These differences may be linked to the loss or absence of one of the optional introns.

Notes: This group sequenced several introns found in cytochrome b oxidase (COX) gens that have open reading frames (some of which were known to be important for maturase activity, others of which were thought to be important for mobility) and noted for the first time the conserved "LAGLIDADG" motif.

1984:

Klar et al. Involvement of double-strand chromosomal breaks for mating-type switching in Saccharomyces cerevisiae. 49: 77 - 88.

The yeast *S. cerevisiae* switches a and alpha cell types by a transposition mechanism that replaces genetic information residing at the mating-type locus (MAT) with information copied from either of the two donor loci, HML and HMR. The donor HML and HMR loci contain the same genetic information as the MAT α and MAT α alleles, yet they do not switch. Additionally, Strathern et al. (1982) have described an in vivo double-strand DNA break found at subgenomic levels (approximately 2% of MAT DNA) within the MAT locus but not within HML and HMR. We have examined the role of this double-strand DNA break in the switching process. Cell lineage studies show that strains containing deletions of the donor HML and HMR loci produce lethal progeny in the exact pattern described for MAT switching in standard strains. Our interpretation is that the double-strand MAT break in the deletion strains cannot be repaired because of the lack of the donor loci, resulting in cell death. We suggest that the double-strand DNA break is an initiating event for switching and that this event is lethal in the absence of the donor loci. MAT mutants isolated as survivors from this "pedigree of death" define a site required for switching where the double-strand break occurs. We have also examined marl mutant strains in which the donor loci are expressed and observed to switch (Klar et al. 1981a). The double-strand DNA cut appears at the HM loci in these strains. Thus, there is a strong correlation between the presence or absence of the double-strand break at each cassette and its ability or inability to switch as observed at the single cell level.

1985:

Jacquier and Dujon, An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene Cell 41: 383- 394 (1985)

The intron of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* possesses a long internal reading frame (ORF) that is conserved in various yeast species. In crosses between intron-plus and intron-minus variants, this intron determines a specific gene conversion phenomenon, which results in the integration of the intron sequence within all previously intron-minus copies of the gene. We show, from a frameshift mutant within the intron ORF and from the need of mitochondrial protein synthesis, that ORF encodes a protein active in the gene conversion that spreads the intron within populations of interbreeding strains. This new intron function is reminiscent of the "transposase" encoded by mobile genetic elements and is discussed in relation to other intron functions.

Zinn and Butow (1985), "Nonreciprocal exchange between alleles of the yeast mitochondrial 21S rRNA gene: kinetics and the involvement of a double-strand break," Cell 40:887-895;

A 1.1 kb intron containing an open reading frame (ORF) in one allele (omega+) of the yeast mitochondrial 21S rRNA gene is nearly quantitatively inserted in crosses into a 21S rRNA allele lacking that intron (omega-). We have determined that this nonreciprocal exchange initiates soon after cells fuse to form zygotes and is complete by 10-16 hr after mating. We have discovered a unique in vivo double-strand cut in omega- mitochondrial DNA (mtDNA) at or near the site of intron insertion that is implicated in the process. Markers flanking the intron insertion site are coconverted with frequencies inversely proportional to their distance from that site. There is no net conversion of omega- to omega+ in crosses between petites retaining these alleles, nor do we observe the unique double-strand cut in the mtDNA from zygotes of such crosses. The data suggest that a translation product of the intron ORF is required for the double-strand cut and nonreciprocal recombination at omega.

Macreadie et al. (1985), "Transposition of an intron in yeast mitochondria requires a protein encoded by that intron," *Cell* 41:395-402)

The optional 1143 bp intron in the yeast mitochondrial 21S rRNA gene (omega +) is nearly quantitatively inserted in genetic crosses into 21S rRNA alleles that lack it (omega -). The intron contains an open reading frame that can encode a protein of 235 amino acids, but no function has been ascribed to this sequence. We previously found an in vivo double-strand break in omega - DNA at or close to the intron insertion site only in zygotes of omega + X omega - crosses that appears with the same kinetics as intron insertion. We now show that mutations in the intron open reading frame that would alter the translation product simultaneously inhibit nonreciprocal omega recombination and the in vivo double-strand break in omega - DNA. These results provide evidence that the open reading frame encodes a protein required for intron transposition and support the role of the double-strand break in the process.

Notes: These three papers in 1985, collectively, demonstrate that the 'polarity of conversion' (the dominant inheritance of the 'ω' group I intron in the 21S rDNA gene in the yeast mitochondrial genome) probably involves: (i) homologous recombination, because a DSB can be observed and because co-conversion of flanking sequence drops off as a function of distance from intron; and (ii) a protein encoded within, and translated from, the group I intron (demonstrated by frame-shift mutations in that ORF that kill intron mobility).

1986:

Colleaux, et al. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* 44: 521 - 533. (1986)

The intron of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* (r1 intron) possesses a 235 codon long internal open reading frame (r1 ORF) whose translation product determines the duplicative transposition of that intron during crosses between intron-plus strains (omega+) and intron-minus ones (omega-). Using site-directed mutagenesis, we have constructed a universal code equivalent of the r1 ORF that, under appropriate promoter control, allows the overexpression in *E. coli* of a protein identical to the mitochondrial intron encoded "transposase". This protein exhibits a double strand endonuclease activity specific for the omega- site. This finding demonstrates, for the first time, the enzymatic activity of an intron encoded protein whose function is to promote the spreading of that intron by generating double strand breaks at a specific sequence within a gene.

Notes: This is the first biochemical characterization of a homing endonuclease using overexpressed protein and a cleavage assay outside of the yeast 21S rDNA target. Authors had to synthesize a codon-optimized gene for *E. coli* translation to do this (not easy in 1985!).

Dujon et al., Mitochondrial introns as mobile genetic elements: the role of intron-encoded proteins. *Basic Life Sci.* 40: 5 - 27.

Introns of organelle genes share distinctive RNA secondary structures that allow their classification into two known families. These structures are believed to play an essential role in splicing, and members of both structural classes have recently been shown to perform self-splicing reactions in vitro. In lower eukaryotes, many structured introns also contain long internal open reading frames (ORFs), which are able to code for hydrophilic proteins. Several properties of self-splicing structured introns suggest that they resemble mobile genetic elements, even though no actual transposition event involving these introns has yet been found. We

report here on the characterization of two intron-encoded proteins that strongly support this attractive idea. First, we show that the class I intron of the 21S ribosomal RNA (rRNA) gene of *Saccharomyces cerevisiae* omega+ strains (rl intron) encodes a specific transposase. This protein has been partially purified from *Escherichia coli* cells that overexpress it from an artificial universal code equivalent to the rl intronic ORF. The omega transposase shows a double-strand endonuclease activity in vitro. This activity creates a 4-bp staggered cut with 3' OH overhangs within a specific sequence of the 21S rRNA gene of omega- strains. It is precisely within this sequence that the rl intron inserts by a duplicative transposition. Second, we report on the synthesis, in *E. coli*, of a putative reverse transcriptase encoded by the class II intron of the cytochrome b gene of *Schizosaccharomyces pombe*. This synthesis was obtained from *E. coli* expression vectors, using the class II intronic ORF linked to an artificial initiator sequence. As further support of the idea that structured introns are mobile, we show, from a systematic screening of introns in various yeast species, that the rl intron has transposed into the ATPase subunit 9 gene of *Kluyveromyces fragilis*. Structural features observed at the new intron homing site may be relevant to the transposition event.

1988:

Colleaux et al. Recognition and cleavage site of the intron-encoded omega transposase. PNAS USA 85: 6022- 6026. (1988).

The optional group I intron of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* contains a 235-codon-long open reading frame the translation product of which (the omega transposase) catalyzes the formation of a double-strand break within the intron-minus (omega-) copies of the same gene. Purified omega transposase generates in vitro a 4-base-pair staggered cut with 3' hydroxyl overhangs at the exact position where the intron eventually inserts in the gene. Using randomly mutagenized synthetic oligonucleotides, single-base mutants were produced at 21 positions around the cleavage site. Experiments with these oligonucleotides show that the recognition site extends over an 18-base pair-long sequence within which minimal sequence degeneracy is tolerated. The intron-encoded omega transposase is, therefore, one of the most specific restriction endonucleases known to date.

Notes: This are the first biochemical studies with purified protein. Results: 18 basepair target, 3' four-base overhangs generated, with intron-insertion site at direct center of cleavage. Degeneracy of recognition with at 13 basepair substitutions out of 29 tested displaying wild-type or only partially reduced cleavage activity, and 16 out of 29 positions displaying complete loss of cleavage activity.

Rudin and Haber Efficient Repair of HO-Induced Chromosomal Breaks in *Saccharomyces cerevisiae* by Recombination between Flanking Homologous Sequences Mol. Cell. Biol. 8 (9): 3918 - 3928. (1988).

Novel recombinational repair of a site-specific double-strand break (DSB) in a yeast chromosome was investigated. When the recognition site for the HO endonuclease enzyme is embedded in nonyeast sequences and placed between two regions of homology, expression of HO endonuclease stimulates recombination between the homologous flanking regions to yield a deletion, the apparent product of an intrachromosomal exchange between direct repeats. This deletion-repair event is very efficient, thus preventing essentially all the potential lethality due to the persistence of a DSB. Interestingly, unlike previous studies involving spontaneous recombination between chromosomal repeats, the recombination events stimulated by HO-induced DSBs are accompanied by loss of the sequences separating the homologous regions greater than 99.5% of the time. Repair is dependent on the RAD52 gene. The deletion-repair event provides an in vivo assay for the sensitivity of any particular recognition site to HO cleavage. By taking advantage of a galactose-inducible HO gene, it has

been possible to follow the kinetics of this event at the DNA level and to search for intermediates in this reaction. Deletion-repair requires approximately 45 min and is inhibited when cycloheximide is added after HO endonuclease cleavage.

Notes: The first study in which recombination in a eukaryotic chromosome (rather than in an extrachromosomal element such as a phage genome or a plasmid) was demonstrated to be induced by a LAGLIDADG protein. The exact structure/function relationship between HO and homing endonucleases such as I-SceI was not known at this point. Although they both are LAGLIDADG enzymes, HO is free-standing (later determined to be descended from an intein-endo fusion) while I-SceI is intron-encoded.

1989:

Quirk et al., Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. Cell 56: 455 - 465 (1989).

Intron mobility in the T-even phages has been demonstrated. Efficient nonreciprocal conversion of intron minus (In-) alleles to intron plus (In+) occurred for the td and sunY genes, but not for nrdB. Conversion to In+ was absolutely dependent on expression of the respective intron open reading frame (ORF). Introns were inserted at their cognate sites in an intronless phage genome via an RNA-independent, DNA-based, duplicative recombination event that was stimulated by exon homology. The td intron ORF product directs the endonucleolytic cleavage of DNA, targeting the site of intron integration. A 21 nucleotide deletion of the integration site abolished high frequency intron inheritance. These experiments provide a novel example of gene conversion in prokaryotes, while suggesting a molecular rationale for the inconsistent distribution of introns within highly conserved exon contexts of the T-even phage genomes.

Notes: This paper described many of the factors influencing intron homing efficiency in clear terms, with a very nice set of experiments. Basically, the authors did a series of studies in which bacteria (harboring a plasmid containing a sequence with an intron + endonuclease + some flanking exon sequence) were infected with phage lacking all of the above, and then the % of phage coming out that had acquired the introns were measured. Using this system, the authors could systematically mutate the intron, the endonuclease ORF, and the exon flanks and see the effect on homing. The authors also reversed the system by having the phage harbor the intron and related sequences and the plasmid act as the recipient, and also demonstrated that the ORF could be provided in trans (as a separate protein-coding construct) and still drive intron mobility.

Delahodde, et al. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria Cell 56: 431 - 441. (1989)

Two introns of the mitochondrial genome 777-3A of *S. cerevisiae*, bl4 in cob and al4 in cox1 genes, contain ORFs that can be translated into two homologous proteins. We changed the UGA, AUA, and CUN codons of these ORFs to the universal genetic code, in order to study the functions of their translated products in *E. coli* and in yeast, by retargeting the nuclear encoded protein into mitochondria. The p27bl4 protein has been shown to be required for the splicing of both introns bl4 and al4. The homologous p28al4 protein is highly toxic to *E. coli*. It can specifically cleave double-stranded DNA at a sequence representing the junction of the two fused flanking exons. We present evidence that this system is a good model for studying the role of mitochondrial intron-encoded proteins in the rearrangement of genetic information at both the RNA (RNA splicing-bl4 maturase) and DNA levels (intron transposition-al4 transposase).

Wenzlau et al., *A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell 10: 421 - 430.*

Some yeast mitochondrial introns encode proteins that promote either splicing (maturases) or intron propagation via gene conversion (the fit1 endonuclease). We surveyed introns in the *cox1* gene for their ability to engage in gene conversion and found that the group I intron, *al4* alpha, was efficiently transmitted to genes lacking it. An endonucleolytic cleavage is detectable in recipient DNA molecules near the site of intron insertion in vivo and in vitro. Conversion is dependent on an intact *al4* alpha open reading frame. This intron product is a latent maturase, but these data show that it is also a potent endonuclease involved in recombination. Dual function proteins that cleave DNA and facilitate RNA splicing may have played a pivotal role in the propagation and tolerance of introns.

Notes: These two papers (Delahodde and Wenzlau) were published back-to-back, and both discuss two introns (*bi4* and *ai4*) and their corresponding intron-encoded proteins (*P-bi4* and *P-ai4*). The basic results are that *P-bi4* is a maturase that catalyzes splicing of both introns, but not an endonuclease; conversely, *P-ai4* is not a maturase but is an endonuclease (that only works on the *ai4* exon junction sequence. One point mutation can restore maturase activity to *P-ia4*.

Dujon (1989), “Group I introns as mobile genetic elements: facts and mechanistic speculations – a review,” *Gene 82:91-114*

Group I introns form a structural and functional group of introns with widespread but irregular distribution among very diverse organisms and genetic systems. Evidence is now accumulating that several group I introns are mobile genetic elements with properties similar to those originally described for the omega system of *Saccharomyces cerevisiae*: mobile group I introns encode sequence-specific double-strand (ds) endoDNases, which recognize and cleave intronless genes to insert a copy of the intron by a ds-break repair mechanism. This mechanism results in: the efficient propagation of group I introns into their cognate sites; their maintenance at the site against spontaneous loss; and, perhaps, their transposition to different sites. The spontaneous loss of group I introns occurs with low frequency by an RNA-mediated mechanism. This mechanism eliminates introns defective for mobility and/or for RNA splicing. Mechanisms of intron acquisition and intron loss must create an equilibrium, which explains the irregular distribution of group I introns in various genetic systems. Furthermore, the observed distribution also predicts that horizontal transfer of intron sequences must occur between unrelated species, using vectors yet to be discovered.

Dujon et al. (1989), “Mobile introns: definition of terms and recommended nomenclature,” *Gene 82 :115-118*

A number of introns in mitochondrial, chloroplast, nuclear or prokaryotic genes have recently been shown to encode double-strand sequence-specific endonucleases. Such introns are mobile genetic elements that insert themselves at or near the cleaved sites. A uniform nomenclature to designate the molecular elements involved in the phenomenon of intron mobility is proposed.

Capecchi *Altering the genome by homologous recombination Science 244: 1288 - 1292. (1989)*

Homologous recombination between DNA sequences residing in the chromosome and newly introduced, cloned DNA sequences (gene targeting) allows the transfer of any modification of the cloned gene into the genome of a living cell. This article discusses the current status of gene targeting with particular emphasis on germ line

modification of the mouse genome, and describes the different methods so far employed to identify those rare embryonic stem cells in which the desired targeting event has occurred.

Notes: this review has nothing to do with homing endonucleases, but is worth reading to understand the state of the art of targeted gene insertions and conversions at the very end of the 1980's, because within about 24 to 36 months (by 1992/1993) investigators were starting to discuss the possibility of using homing endonucleases to target homologous recombination / integration to desired loci.

1990:

Bell-Pederson, et al. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nuc. Acids Res. 18: 3763 - 3770 (1990).

Although mobility of the phylogenetically widespread group I introns appears to be mechanistically similar, the phage T4 intron-encoded endonucleases that promote mobility of the td and sunY introns are different from their eukaryotic counterparts. Most notably, they cleave at a distance from the intron insertion sites. The td enzyme was shown to cleave 23-26 nt 5' and the sunY endonuclease 13-15 nt 3' to the intron insertion site to generate 3-nt or 2-nt 3'-OH extensions, respectively. The absolute coconversion of exon markers between the distant cleavage and insertion sites is consistent with the double-strand-break repair model for intron mobility. As a further critical test of the model we have demonstrated that the mobility event is independent of DNA sequences that encode the catalytic intron core structure. Thus, in derivatives in which the lacZ or kanR coding sequences replace the intron, these marker genes are efficiently inserted into intron-minus alleles when the cognate endonuclease is provided in trans. The process is therefore endonuclease-dependent, rather than dependent on the intron per se. These findings imply that the endonucleases rather than the introns themselves were the primordial mobile elements and are incorporated into a model for the evolution of mobile introns.

Notes: This is a truly crucial paper in the field. All previous studies of homing demonstrated the ability of the homing endonucleases to facilitate the transfer of their cognate intron to intron-minus hosts, presumably via homologous recombination. Those studies did not preclude the possibility that the intron itself might participate in intermediate species formed during homing. In this experiment, Bell-Pederson et al. demonstrated that they could entirely replace the intron with an unrelated selectable marker (kanR) and by supplying the endonuclease in trans (expressed separately) homing of the KanR gene could be induced into a site containing the endonuclease target, and homology to the sequences flanking the KanR marker. Thus, homing only requires homology between donor and acceptor, the endonuclease and its target sequence in the acceptor DNA.

Muscarella et al., Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of Physarum polycephalum. Mol. Cell Biol. 10:3386-96 (1990).

A novel class of enzymes is composed of the site-specific endonucleases encoded by some group I introns. We have characterized several aspects of I-Ppo, the endonuclease that mediates the mobility of intron 3 in the ribosomal DNA of *Physarum polycephalum*. This intron is unique among mobile group I introns in that it is located in nuclear DNA. We found that I-Ppo is encoded by an open reading frame in the 5' half of intron 3, upstream of the sequences required for self-splicing of group I introns. Either of two AUG initiation codons could start this reading frame. The longer polypeptide was the major form translated in vitro in a reticulocyte extract. From nuclease assays of proteins synthesized in vitro with partially deleted DNAs, we conclude that both polypeptides possess endonuclease activity. We also have expressed I-Ppo in *Escherichia coli*, using a bacteriophage T7 RNA polymerase expression system. The longer polypeptide also was the predominant form made in this system. It showed enzymatic activity in bacteria in vivo, as demonstrated by the cleavage of a plasmid carrying the target site. Like several other intron-encoded endonucleases, I-Ppo makes a four-base staggered cut in its ribosomal DNA target sequence, very near the site where intron 3 becomes integrated in crosses of intron 3-containing and intron 3-lacking *Physarum* strains.

Notes: This is another important paper. First, this paper describes the discovery of the 'His-Cys' box family of homing endonucleases (separate from the LAGLIDADG and the GIY-YIG families; distantly related to the HNH families). More importantly, this is the first example of a mobile intron and homing endonuclease, and accompanying homing activity, observed in a nuclear genome (where DNA condensation and chromatin packaging might be expected to interfere with endonuclease function).

However, it is worth noting that rDNA genes in *Physarum* are actually extrachromosomal elements, localized to the nucleolus, and do not package and condense like normal chromosomes:

"rDNA molecules in *Physarum* are confined to the nucleolus in interphase. Our study definitively locates these extrachromosomal genes in mitosis in the form of thin DNA fibers contained within nucleolar remnants. We further show that these rDNA minichromosomes do not condense and that they segregate as entities independent of the condensed chromosomal DNA. Our results illustrate the discontinuous nature of the nucleolar organizing region in *Physarum*." *Dutilleule and Pierron, Localization by high resolution in situ hybridization of the ribosomal minichromosomes during the nucleolar cycle of Physarum polycephalum Exp. Cell. Reseach 203 (2): 354 - 364.*

Montheilet et al. Purification and characterization of the in vitro activity of I-SceI, a novel and highly specific endonuclease encoded by a group I intron Nuc. Acids Res. 18: 1407 - 1411 (1990).

Group I intron encoded proteins represent a novel class of site specific double strand endonucleases. The endonuclease activity of this class of proteins has been first demonstrated In vivo for I-SceI which is encoded by a mitochondrial intron of *Saccharomyces cerevisiae*. Assays using crude cell extracts have shown that I-Sce I can be used in vitro as a restriction endonuclease potentially useful for recombinant DNA technology owing to its large recognition sequence (18 nucleotides). We report here the purification and the first detailed analysis of the In vitro activity and properties of I-Sce I.

Sargueil et al. (1990) *In vivo and in vitro analyses of an intron-encoded endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis. Nuc. Acids Res. 18 (19): 5659 - 5664.*

The pal 4 nuclease (termed I-SceII) is encoded in the group I al 4 intron of the COX I gene of *Saccharomyces cerevisiae*. It introduces a specific double-strand break at the junction of the two exons A4-A5 and thus mediates the insertion of the intron into an intronless strain. To define the sequence recognized by pal 4 we introduced 35 single mutations in its target sequence and examined their cleavage properties either *in vivo* in *E. coli* (when different forms of the pal 4 proteins were artificially produced) or *in vitro* with mitochondrial extracts of a mutant yeast strain blocked in the splicing of the al 4 intron. We also detected the pal 4 DNA endonuclease activity in extracts of the wild type strain. The results suggest that 6 to 9 noncontiguous bases in the 17 base-pair region examined are necessary for pal 4 nuclease to bind and cleave its recognition site. We observed that the pal 4 nuclease specificity can be significantly different with the different forms of the protein thus explaining why only some forms are highly toxic in *E. coli*. This study shows that pal 4 recognition site is a complex phenomenon and this might have evolutionary implications on the transfer properties of the intron.

Nickoloff et al. *In Vivo Analysis of the Saccharomyces cerevisiae HO Nuclease Recognition Site by Site-Directed Mutagenesis Mol. Cell. Biol. 10: 1174 - 1179.*

HO nuclease introduces a specific double-strand break in the mating-type locus (MAT) of *Saccharomyces cerevisiae*, initiating mating-type interconversion. To define the sequence recognized by HO nuclease, random mutations were produced in a 30-base-pair region homologous to either MATot or MATa by a chemical synthesis procedure. The mutant sites were introduced into *S. cerevisiae* on a shuttle vector and tested for the ability to stimulate recombination in an assay that mimics mating-type interconversion. The results suggest that a core of 8 noncontiguous bases near the Y-Z junction of MAT is essential for HO nuclease to bind and cleave its recognition site. Other contacts must be required because substrates that contain several mutations outside an intact core reduce or eliminate cleavage *in vivo*. The results show that HO site recognition is a complex phenomenon, similar to promoter-polymerase interactions.

Notes: These three papers are among the first that really start to work out details of activity and specificity profiles for LAGLIDADG endonucleases.

1991:

Durrenberger and Rochaix, *Chloroplast ribosomal intron of Chlamydomonas reinhardtii: in vitro self-splicing, DNA endonuclease activity and in vivo mobility. EMBO Journal 10: 3495 - 3501. (1991)*

All chloroplast 23S ribosomal RNA genes of the unicellular alga *Chlamydomonas reinhardtii* contain an 888 bp group I intron with an internal open reading frame (ORF). A precursor RNA encompassing the intron with its 5' and 3' flanking sequences was shown to self-splice both during *in vitro* transcription and upon incubation of the isolated pre-RNA under self-splicing conditions. Expression of the internal ORF in *Escherichia coli* in the presence of a plasmid containing a cDNA corresponding to the intronless form of the 23S rRNA gene resulted in specific cleavage of the cDNA at or close to the exon junction sequence. To test whether this ORF-encoded double-strand DNA endonuclease is involved in intron mobility *in vivo*, the same ribosomal cDNA was stably integrated into the *C. reinhardtii* chloroplast genome using particle gun mediated transformation. All the transformants with the cDNA integrated at the expected site in the chloroplast genome had the intron precisely inserted at the artificial exon junction site. These experiments demonstrate that the chloroplast ribosomal intron

of *C. reinhardtii* behaves as a ribozyme in vitro and also as a mobile genetic element in vivo provided a target site is present.

Notes: This paper describes the discovery of I-CreI--the hallmark enzyme of the LAGLIDADG homodimer subfamily, that has subsequently been adopted many times for the creation of gene targeting homing endonucleases.

Bell-Pedersen et al., I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target PNAS 88:7719-7723 (1991).

Mobility of the phage T4 td intron depends on activity of an intron-encoded endonuclease (I-TevI), which cleaves a homologous intronless (Δ In) target gene. The double-strand break initiates a recombination event that leads to intron transfer. We found previously that I-TevI cleaves Δ In target DNA 23-26 nucleotides upstream of the intron insertion site. DNase I-footprinting experiments and gel-shift assays indicate that I-TevI makes primary contacts around the intron insertion site. A synthetic DNA duplex spanning the insertion site but lacking the cleavage site was shown to bind I-TevI specifically, and when cloned, to direct cleavage into vector sequences. The behavior of the cloned duplex and that of deletion and insertion mutants support a primary role for sequences surrounding the insertion site in directing I-TevI binding, conferring cleavage ability, and determining cleavage polarity. On the other hand, sequences around the cleavage site were shown to influence cleavage efficiency and cut-site selection. The role of cleavage-site sequences in determining cleavage distance argues against a strict "ruler" mechanism for cleavage by I-TevI. The complex nature of the homing site recognized by this unusual type of endonuclease is considered in the context of intron spread.

Marshall & Lemieux, Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast algae subunit rRNA-encoding gene of Chlamydomonas eugametos Gene 15:241-5 (1991).

The fifth group-I intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos* (CeLSU.5) is mobile during interspecific crosses between *C. eugametos* and *Chlamydomonas moewusii*. Like the six other mobile introns that have been well characterized so far, CeLSU.5 contains a long open reading frame (ceuIR) coding for a site-specific endonuclease (I-CeuI) that cleaves the *C. moewusii* intronless gene in the vicinity of the intron-insertion site. This stimulates gap repair and mediates efficient transfer of the intron at its cognate site. By expressing the ceuIR gene in the *Escherichia coli* vectors pKK233-2 and pTRC-99A, we recently demonstrated that the endonuclease is highly toxic to *E. coli* [Gauthier et al., *Curr. Genet.* 19 (1991) 43-47]. To eliminate this problem and characterize the cleavage pattern and recognition sequence of the I-CeuI endonuclease, we have expressed the ceuIR gene in *E. coli* under the control of a bacteriophage T7 promoter in a tightly regulated M13 system, and developed an in vitro system to assay partially purified I-CeuI activity. This allowed us to determine that I-CeuI recognizes a sequence of less than 26 bp centered around the insertion site and produces a staggered cut 5 bp downstream from this site, yielding 4-nucleotide (CTAA), 3'-OH overhangs.

Sargueil, A new specific DNA endonuclease activity in yeast mitochondria. Mol. Gen Gen 225:340-341 (1991).

Two group I intron-encoded proteins from the yeast mitochondrial genome have already been shown to have a specific DNA endonuclease activity. This activity mediates intron insertion by cleaving the DNA sequence corresponding to the splice junction of an intronless strain. We have discovered in mitochondrial extracts from the yeast strain 777-3A a new DNA endonuclease activity which cleaves the fused exon A3-exon A4 junction sequence of the COXI gene.

Kawasaki et al., DNA Sequence Recognition by a Eukaryotic Sequence-Specific Endonuclease, Endo.SceI, from Saccharomyces cerevisiae. J. Biol. Chem. 266:5342-5347 (1991).

A eukaryotic sequence-specific endonuclease, Endo.SceI, causes sequence-specific double-stranded scission of double-stranded DNA to produce cohesive ends with four bases protruding at the 3' termini. Unlike in the case of restriction enzymes, an asymmetric 26-base pair consensus sequence was found around the cleavage site for Endo.SceI instead of a common sequence. We analyzed the base pairs that interacted with Endo.SceI on the recognition of its cleavage sites. A region comprising -10 through +16 base pairs from the center of the cleavage site was shown to be essential and sufficient for the sequence-specific cutting with Endo.SceI by experiments involving synthesized DNAs. Methylation interference experiments indicate that bases in the region comprising the +7 through +14 base pairs is involved in close contact with Endo.SceI in its recognition of the cleavage site. This +7 through +14-base pair region overlaps the most stringently conserved sequence in the consensus sequence for the cleavage site, suggesting that this region constitutes the core for the recognition by Endo.SceI.

Eddy and Gold, The phage T4 nrdB intron: a deletion mutant of a version found in the wild Genes Dev. 5: 1032 - 1041.

Bacteriophage T4 possesses three self-splicing group I introns. Two of the three introns are mobile elements; the third, in the gene encoding a subunit of the phage nucleotide reductase (nrdB), is not mobile. Because intron mobility offers a reasonable explanation for the paradoxical occurrence of large intervening sequences in a space-efficient eubacterial phage, it is puzzling that the nrdB intron is not mobile like its compatriots. We have discovered a larger nrdB intron in a closely related phage, and we infer from comparative sequence data that the T4 intron is a deletion mutant derived from this larger intron. This larger nrdB intron encodes an open reading frame of 269 codons, which we have cloned and overexpressed. The overexpressed protein shows a dsDNA endonuclease activity specific for the intronless nrdB gene, typical of mobile introns. Thus, we believe that all three introns of T4 are or were mobile "infectious introns" and that they have entered into and been maintained in the phage population by virtue of this efficient mobility.

Schiestl and Petes, Integration of DNA fragments by illegitimate recombination in Saccharomyces cerevisiae PNAS USA 88: 7585 - 7589.

DNA fragments (generated by BamHI treatment) with no homology to the yeast genome were transformed into *Saccharomyces cerevisiae*. When the fragments were transformed in the presence of the BamHI enzyme, they integrated into genomic BamHI sites. When the fragments were transformed in the absence of the enzyme, they integrated into genomic G-A-T-C sites. Since the G-A-T-C sequence is present at the ends of BamHI fragments, this results indicates that four base pairs of homology are sufficient for some types of mitotic recombination.

1992:

Ma et al., The group I intron of apocytochrome b gene from Chlamydomonas smithii encodes a site-specific endonuclease. Plant Mol. Biol. 18:1001-1004 (1992).

The mitochondrial DNA molecules of two interfertile algal species, *Chlamydomonas smithii* and *C. reinhardtii*, are co-linear except for a 1075 bp intron (the alpha-insert) that is present in the cob gene of *C. smithii*. The alpha-insert, a group I intron (*Cs cob.1*) containing an open reading frame (ORF) which encodes a basic,

hydrophilic protein of 237 amino acids, is unidirectionally transmitted to all diploid progeny during interspecific crosses. In this report, we show that the Cs cob.1-encoded protein is a site-specific endonuclease (I-Csm I) which could mediate the intron transfer via the gene conversion mechanism. The Cs cob.1 ORF was cloned into the vector pMALcr1 and over-expressed as a hybrid protein fused to maltose-binding protein (MBP). This fusion protein exhibited an in vivo endonuclease activity which specifically cleaved the intron homing site within the intronless cob gene.

Sharma et al., Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proc. Natl. Acad. Sci. USA 89:6658-62 (1992).

The bacteriophage T4 segA gene lies in a genetically unmapped region between the gene beta gt (beta-glucosyltransferase) and uvsX (recombination protein) and encodes a protein of 221 amino acids. We have found that the first 100 amino acids of the SegA protein are highly similar to the N termini of four other predicted T4 proteins, also of unknown function. Together these five proteins, SegA-E (similar to endonucleases of group I introns), contain regions of similarity to the endonuclease I-Tev I, which is encoded by the mobile group I intron of the T4 td gene, and to putative endonucleases of group I introns present in the mitochondria of Neurospora crassa, Podospora anserina, and Saccharomyces douglasii. Intron-encoded endonucleases are required for the movement (homing) of the intron DNA into an intronless gene, cutting at or near the site of intron insertion. Our in vitro assays indicate that SegA, like I-Tev I, is a Mg(2+)-dependent DNA endonuclease that has preferred sites for cutting. Unlike the I-Tev I gene, however, there is no evidence that segA (or the other seg genes) resides within introns. Thus, it is possible that segA encodes an endonuclease that is involved in the movement of the endonuclease-encoding DNA rather than in the homing of an intron.

Plessis et al. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics 130: 451 - 460. (1992). Submitted Sept. 20; accepted Nov. 22.

The Saccharomyces cerevisiae mitochondrial endonuclease I-SceI creates a double-strand break as the initiating step in the gene conversional transfer of the omega+ intron to omega- DNA. We have expressed a galactose-inducible synthetic I-SceI gene in the nucleus of yeast that also carries the I-SceI recognition site on a plasmid substrate. We find that the galactose-induced I-SceI protein can be active in the nucleus and efficiently catalyze recombination. With a target plasmid containing direct repeats of the Escherichia coli lacZ gene, one copy of which is interrupted by a 24-bp cutting site, galactose induction produces both deletions and gene conversions. Both the kinetics and the proportion of deletions and gene conversions are very similar to analogous events initiated by a galactose-inducible HO endonuclease gene. We also find that, in a rad52 mutant strain, the repair of double-strand breaks initiated by I-SceI and by HO are similarly affected: the formation of deletions is reduced, but not eliminated. Altogether, these results suggest either that the two endonucleases act in the same way after double-strand break formation or that the two endonucleases are not involved in subsequent steps.

Wernette et al., Complex recognition site for the group I intron-encoded endonuclease I-SceII Mol. Cell Biol. 12:1903 (1992).

We have characterized features of the site recognized by a double-stranded DNA endonuclease, I-SceII, encoded by intron 4 alpha of the yeast mitochondrial COX1 gene. We determined the effects of 36 point mutations on the cleavage efficiency of natural and synthetic substrates containing the Saccharomyces capensis I-SceII site. Most mutations of the 18-bp I-SceII recognition site are tolerated by the enzyme, and those mutant sites are cleaved between 42 and 100% as well as the wild-type substrate is. Nine mutants blocked cleavage to less than or equal to 33% of the wild-type, whereas only three point mutations, G-4----C, G-12----T, and G-15-- --C, block cleavage completely. Competition experiments indicate that these three substrates are not cleaved, at

least in part because of a marked reduction in the affinity of the enzyme for those mutant DNAs. About 90% of the DNAs derived from randomization of the nucleotide sequence of the 4-bp staggered I-SceII cleavage site are not cleaved by the enzyme. I-SceII cleaves cloned DNA derived from human chromosome 3 about once every 110 kbp. The I-SceII recognition sites in four randomly chosen human DNA clones have 56 to 78% identity with the 18-bp site in yeast mitochondrial DNA; they are cleaved at least 50% as well as the wild-type mitochondrial substrate despite the presence of some substitutions that individually compromise cleavage of the mitochondrial substrate. Analysis of these data suggests that the effect of a given base substitution in I-SceII cleavage may depend on the sequence at other positions.

Seraphin et al., The yeast mitochondrial intron aI5 alpha; associated endonuclease activity and in vivo mobility Gene 113:1-8 (1992).

By analyzing crosses between yeast strains carrying different combinations of mitochondrial (mt) introns, we have shown that the aI5 alpha intron is mobile in vivo. Furthermore, we have observed that the mobility of intron aI5 alpha is affected by both the nuclear and mt genotypes. We have also detected a restriction endonuclease (ENase) activity that cleaves intronless mt genomes close to the aI5 alpha intron insertion site and thus might be involved in intron mobility. This is further supported by the fact that this ENase activity is only detected in a strain containing the aI5 alpha intron. Furthermore, similar to other ENases encoded by mobile mt introns of yeast, the ENase generates a cut with a four-base 3'-OH overhang. Thus, intron aI5 alpha represents a characteristic member of the family of mobile group-I introns.

Marshall and Lemieux The I-CeuI endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the Chlamydomonas moewusii chloroplast large subunit rRNA gene. Nuc. Acids Res. 20: 6401 - 6407. (1992).

The I-CeuI endonuclease is a member of the growing family of homing endonucleases that catalyse mobility of group I introns by making a double-strand break at the homing site of these introns in cognate intronless alleles during genetic crosses. In a previous study, we have shown that a short DNA fragment of 26 bp, encompassing the homing site of the fifth intron in the Chlamydomonas eugametos chloroplast large subunit rRNA gene (Ce LSU.5), was sufficient for I-CeuI recognition and cleavage. Here, we report the recognition sequence of the I-CeuI endonuclease, as determined by random mutagenesis of nucleotide positions adjacent to the I-CeuI cleavage site. Single-base substitutions that completely abolish endonuclease activity delimit a 15-bp sequence whereas those that reduce the cleavage rate define a 19-bp sequence that extends from position -7 to position +12 with respect to the Ce LSU.5 intron insertion site. As the other homing endonucleases that have been studied so far, the I-CeuI endonuclease recognizes a non-symmetric degenerate sequence. The top strand of the recognition sequence is preferred for I-CeuI cleavage and the bottom strand most likely determines the rate of double-strand breaks.

Frey et al., Specific cleavage of the yeast and bacterial genomes at a single site using the rare cutter endonuclease I-SceI (Megannuclease I-SceI) Fresenius J. Anal.Chem. 343: 122 - 123.

1993:

Puchta et al. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nuc. Acids Res. 21: 5034 - 5040.

Induction of double strand breaks (DSBs) is coupled to meiotic and mitotic recombination in yeast. We show that also in a higher eukaryote induction of DSBs is directly correlated with a strong enhancement of recombination frequencies. We cotransfected *Nicotiana plumbaginifolia* protoplasts with a plasmid carrying a synthetic I-SceI gene, coding for a highly sequence specific endonuclease, together with recombination substrates carrying an I-SceI-site adjacent to their homologous sequences. We measured efficiencies of extrachromosomal recombination, using a well established transient beta-glucuronidase (GUS) assay. GUS enzyme activities were strongly increased when a plasmid carrying the I-SceI gene in sense but not in antisense orientation with respect to the promoter was included in the transfections. The in vivo induced DSBs were detected in the recombination substrates by Southern blotting, demonstrating that the yeast enzyme is functional in plant cells. At high ratios of transfected I-SceI-genes to I-SceI-sites the majority of the I-SceI-sites in the recombination substrates are cleaved, indicating that the induction of the DSBs is the rate limiting step in the described recombination reaction. These results imply that in vivo induction of transient breaks at specific sites in the plant genome could allow foreign DNA to be targeted to these sites via homologous recombination.

Perrin et al. (1993), "Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions," EMBO J. 12:2939-2947

Group I intron-encoded endonucleases represent a new class of double strand cutting endonucleases whose function is to initiate the homing of introns by generating double strand breaks in site-specific sequences. We have studied the mechanism of interaction of the I-SceI endonuclease with different DNA substrates derived from its natural site in the intron-less gene or from intron-exon junctions in the gene with an intron. We show that the enzyme recognizes its asymmetrical site with high affinity binding to the sequence corresponding to the downstream exon followed by binding to the upstream exon and catalysis of phosphodiester bond hydrolysis. Asymmetrical nicking activity is observed as an intermediate of the cleavage reaction. In the intron-containing gene, the enzyme recognizes the downstream intron-exon junction without any cleavage activity. This binding raises the possibility of a specific function of homing endonucleases in either gene expression or intron homing steps subsequent to DNA cleavage.

1994:

Rouet et al., Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Mol. Cell Biol. 14: 8096 – 8106 (1994).

Double-strand breaks introduced into DNA in vivo have been shown to enhance homologous recombination in a variety of chromosomal and extrachromosomal loci in *Saccharomyces cerevisiae*. To introduce double-strand breaks in DNA at defined locations in mammalian cells, we have constructed a mammalian expression vector for a modified form of I-Sce I, a yeast mitochondrial intron-encoded endonuclease with an 18-bp recognition sequence. Expression of the modified I-Sce I endonuclease in COS1 cells results in cleavage of model recombination substrates and enhanced extrachromosomal recombination, as assayed by chloramphenicol acetyltransferase activity and Southern blot analysis. Constitutive expression of the endonuclease in mouse 3T3 cells is not lethal, possibly due to either the lack of I-Sce I sites in the genome or sufficient repair of them. Expression of an endonuclease with such a long recognition sequence will provide a powerful approach to studying a number of molecular processes in mammalian cells, including homologous recombination.

Rouet et al. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. 14: 8096 - 8116.

To maintain genomic integrity, double-strand breaks (DSBs) in chromosomal DNA must be repaired. In mammalian systems, the analysis of the repair of chromosomal DSBs has been limited by the inability to introduce well-defined DSBs in genomic DNA. In this study, we created specific DSBs in mouse chromosomes for the first time, using an expression system for a rare-cutting endonuclease, I-SceI. A genetic assay has been devised to monitor the repair of DSBs, whereby cleavage sites for I-SceI have been integrated into the mouse genome in two tandem neomycin phosphotransferase genes. We find that cleavage of the I-SceI sites is very efficient, with at least 12% of stably transfected cells having at least one cleavage event and, of these, more than 70% have undergone cleavage at both I-SceI sites. Cleavage of both sites in a fraction of clones deletes 3.8 kb of intervening chromosomal sequences. We find that the DSBs are repaired by both homologous and nonhomologous mechanisms. Nonhomologous repair events frequently result in small deletions after rejoining of the two DNA ends. Some of these appear to occur by simple blunt-ended ligation, whereas several others may occur through annealing of short regions of terminal homology. The DSBs are apparently recombinogenic, stimulating gene targeting of a homologous fragment by more than 2 orders of magnitude. Whereas gene-targeted clones are nearly undetectable without endonuclease expression, they represent approximately 10% of cells transfected with the I-SceI expression vector. Gene targeted clones are of two major types, those that occur by two-sided homologous recombination with the homologous fragment and those that occur by one-sided homologous recombination. Our results are expected to impact a number of areas in the study of mammalian genome dynamics, including the analysis of the repair of DSBs and homologous recombination and, potentially, molecular genetic analyses of mammalian genomes.

1995:

Smih et al. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells Nuc. Acids Research 23: 5012 – 5019 (1995).

Double-strand breaks (DSBs) are recombinogenic lesions in chromosomal DNA in yeast, *Drosophila* and *Caenorhabditis elegans*. Recent studies in mammalian cells utilizing the I-SceI endonuclease have demonstrated that in some immortalized cell lines DSBs in chromosomal DNA are also recombinogenic. We have now tested embryonic stem (ES) cells, a non-transformed mouse cell line frequently used in gene targeting studies. We find that a DSB introduced by I-SceI stimulates gene targeting at a selectable neo locus at least 50-fold. The enhanced level of targeting is achieved by transient expression of the I-SceI endonuclease. In 97% of targeted clones a single base pair polymorphism in the transfected homologous fragment was incorporated into the target locus. Analysis of the targeted locus demonstrated that most of the homologous recombination events were 'two-sided', in contrast to previous studies in 3T3 cells in which 'one-sided' homologous events predominated. Thus ES cells may be more faithful in incorporating homologous fragments into their genome than other cells in culture.

Choulika et al. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 15: 1968 – 1973 (1995).

The mitochondrial intron-encoded endonuclease I-SceI of *Saccharomyces cerevisiae* has an 18-bp recognition sequence and, therefore, has a very low probability of cutting DNA, even within large genomes. We demonstrate that double-strand breaks can be initiated by the I-SceI endonuclease at a predetermined location in the mouse genome and that the breaks can be repaired with a donor molecule homologous regions flanking the breaks. This induced homologous recombination is approximately 2 orders of magnitude more frequent than spontaneous homologous recombination and at least 10 times more frequent than random integration near an active promoter. As a consequence of induced homologous recombination, a heterologous novel sequence can

be inserted at the site of the break. This recombination can occur at a variety of chromosomal targets in differentiated and multipotential cells. These results demonstrate homologous recombination involving chromosomal DNA by the double-strand break repair mechanism in mammals and show the usefulness of very rare cutter endonucleases, such as I-SceI, for designing genome rearrangements.