# Genetic Analysis of the *Chlamydomonas reinhardtii* I-*Cre*I Mobile Intron Homing System in *Escherichia coli*

**Lenny M. Seligman,*,† Kathryn M. Stephens,* Jeremiah H. Savage† and Raymond J. Monnat, Jr.***

**Department of Pathology, University of Washington, Seattle, Washington 98195-7705 and*
*†Department of Biology, Pomona College, Claremont, California 91711*

Manuscript received May 22, 1997
Accepted for publication September 17, 1997

ABSTRACT

We have developed and used a genetic selection system in *Escherichia coli* to study functional requirements for homing site recognition and cleavage by a representative eukaryotic mobile intron endonuclease. The homing endonuclease, I-*Cre*I, was originally isolated from the chloroplast of the unicellular green alga *Chlamydomonas reinhardtii*. I-*Cre*I homing site mutants contained base pair substitutions or single base deletions that altered the rate of homing site cleavage and/or product release. I-*Cre*I endonuclease mutants fell into six phenotypic classes that differed in *in vivo* activity, toxicity or genetic dominance. Inactivating mutations clustered in the N-terminal 60% of the I-*Cre*I amino acid sequence, and two frameshift mutations were isolated that resulted in premature translation termination though retained partial activity. These mutations indicate that the N-terminal two-thirds of the I-*Cre*I endonuclease is sufficient for homing site recognition and cleavage. Substitution mutations altered in four potential active site residues were examined: D20N, Q47H or R70A substitutions inactivated endonuclease activity, whereas S22A did not. The genetic approach we have taken complements phylogenetic and structural studies of mobile intron endonucleases and has provided new information on the mechanistic basis of I-*Cre*I homing site recognition and cleavage.

MOBILE intron endonucleases are a diverse family of proteins encoded by introns that can self-splice at the RNA level or, as protein introns ("inteins"), at the protein level. The endonucleolytic activity of these proteins can promote lateral transfer (or "homing") of their encoding introns to specific sites in complex genomes. The intron homing cycle (Figure 1A) is initiated and targeted by a site-specific DNA double strand break made by the intron-encoded endonuclease in a homing site allele that lacks the intron. A copy of the intron containing the endonuclease open reading frame is transferred to the cleaved homing site by DNA double strand break repair or gene conversion (LAMBOWITZ and BELFORT 1993; MUELLER *et al.* 1993; BELFORT and PERLMAN 1995; BELFORT and ROBERTS 1997). The inserted intron is replicated with host DNA and expresses the intron-encoded endonuclease, thus retaining the potential for homing to other endonuclease-sensitive, intron-minus alleles.

Mobile introns encoding endonucleases have been identified in organelles, and in a few instances the nuclei, of many different unicellular eucaryotes. They have also been identified in eubacteria, in *Archea* and in bacteriophage genomes. Homing of these encoding introns occurs in the context of mating or transformation between intron-plus and intron-minus strains or, in the case of bacteriophage mobile introns, in the context of infection. Mobile intron endonuclease homing sites differ markedly from the recognition and cleavage sites of other site-specific endonucleases. For example, the type II restriction endonucleases recognize and cleave short (4–8 bp) DNA sequences that are often perfectly twofold sequence-symmetric or palindromic (ROBERTS and HALFORD 1993). Mobile intron homing sites, in contrast, are considerably longer (15–40 bp), and often display little or no twofold sequence symmetry (LAMBOWITZ and BELFORT 1993; MUELLER *et al.* 1993; BELFORT and ROBERTS 1997).

The length and specificity of cleavage of many homing sites, their location in physiologically important host genes and the ability of many mobile introns to self-splice at the RNA or protein level suggest that intron homing may have arisen as a form of "smart" DNA parasitism: once present within an essential gene, a mobile intron can self-splice at the RNA or protein level to minimize deleterious effects on the host and thus insure intron propagation with the potential for lateral transfer (reviewed in LAMBOWITZ and BELFORT 1993; MUELLER *et al.* 1993; BELFORT *et al.* 1995; BELFORT and ROBERTS 1997).

We have developed and used a genetic system in *Escherichia coli* to study functional requirements for homing site recognition and cleavage by I-*Cre*I, a eucaryotic mobile intron endonuclease. The 888-bp I-*Cre*I mobile in-

*Corresponding author:* Raymond J. Monnat, Jr., Department of Pathology, University of Washington, Box 357705, Seattle, WA 98195-7705. E-mail: monnat@u.washington.edu

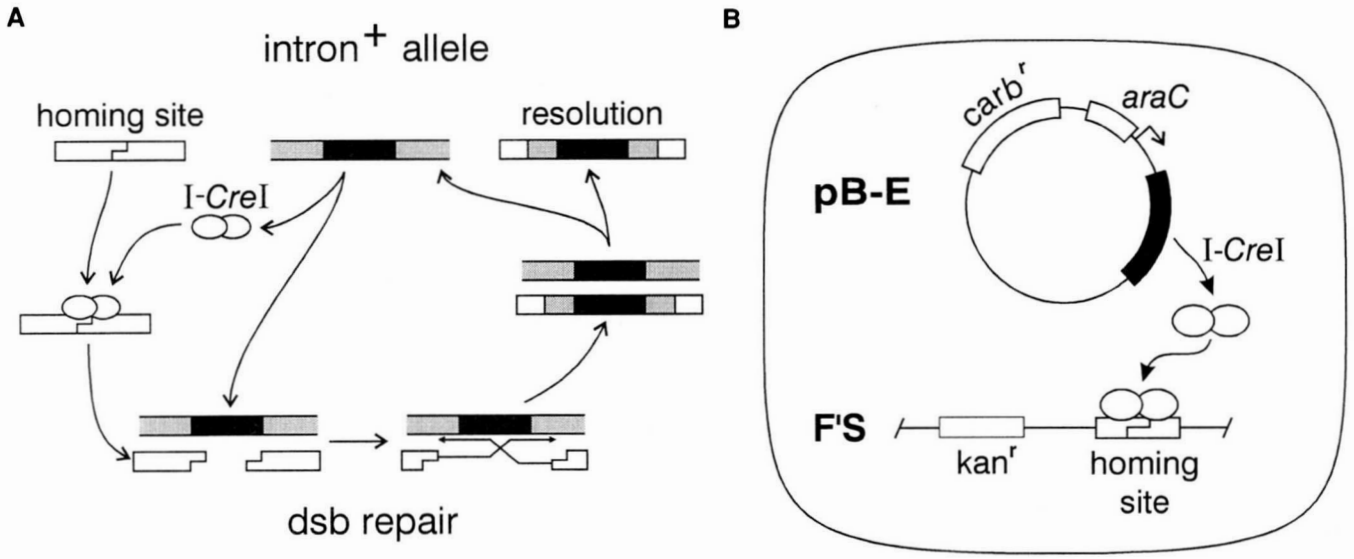


FIGURE 1.—I-*Cre*I mobile intron homing cycle and genetic assay for I-*Cre*I homing site recognition and cleavage in *E. coli*. (A) Homing of the I-*Cre*I mobile intron is initiated by a site-specific DNA double strand break made by I-*Cre*I endonuclease encoded by the mobile intron open reading frame (filled box of intron$^+$ allele, at top center). The I-*Cre*I homing site is located in the *C. reinhardtii* chloroplast 23S rRNA gene (open box, top left). The cleaved homing site is a substrate for double strand break repair off the intron-plus allele. Repair transfers a copy of the intron to the cleaved recipient allele (dsb repair, bottom center). Resolution (resolution, top right) generates two alleles containing the mobile intron, each flanked by a disrupted I-*Cre*I homing site. (B) A genetic assay for I-*Cre*I endonuclease function was established in *E. coli* by placing the I-*Cre*I endonuclease open reading frame on a plasmid, pB-E, under the control of arabinose operon regulatory sequences (the *ara* promoter, $P^*$, and *araC* regulator) and an I-*Cre*I homing site on an F128 F′ factor adjacent to a kanamycin resistance gene (F′S). Expression of active I-*Cre*I endonuclease leads to the loss of the F′ kanamycin marker, presumably due to homing site cleavage. Mutations that inactivate either endonuclease function or that render the homing site resistant to cleavage generate cells that retain kanamycin resistance. Homing with intron transfer does not occur in this assay due to lack of the 888-bp mobile intron (see A).

tron was originally identified in the chloroplast genome of the unicellular green alga *Chlamydomonas reinhardtii* (ROCHAIX and MALNOE 1978; ROCHAIX *et al.* 1985; LEMIEUX *et al.* 1989). This intron can self-splice at the RNA level and can home to copies of the 23S rRNA gene that lack the mobile intron (HERRIN *et al.* 1990; DÜRRENBERGER and ROCHAIX 1991; THOMPSON and HERRIN 1991; DÜRRENBERGER *et al.* 1996).

The 163-residue I-*Cre*I endonuclease protein is encoded by a 489-bp open reading frame within the I-*Cre*I mobile intron. The I-*Cre*I recognition and cleavage site is a 19–24-bp sequence in the 3′ end of the *C. reinhardtii* chloroplast 23S rRNA gene. This site displays limited, twofold sequence symmetry around four central basepairs that are asymmetrically cleaved to generate four base, 3′OH-extended cohesive ends. Intron insertion occurs 1 bp 5′ to the cleaved central four basepairs of the I-*Cre*I homing site, separating the halves of the homing site to generate an intron-plus allele that is endonuclease-resistant (ROCHAIX *et al.* 1985; LEMIEUX *et al.* 1989; THOMPSON *et al.* 1992; DÜRRENBERGER and ROCHAIX 1993; see Figure 2).

Our aim in initiating a genetic analysis of the I-*Cre*I homing system was to develop a facile genetic system to determine sequence requirements for I-*Cre*I intron homing site recognition and cleavage. We reasoned that a genetic analysis of endonuclease and homing site requirements for homing site recognition and cleavage, in conjunction with recent biochemical data and the X-ray crystal structure of native I-*Cre*I protein (HEATH *et al.* 1997; STEPHENS *et al.* 1997; K. M. STEPHENS, unpublished results), would begin to provide a detailed picture of the structural and mechanistic bases for homing site recognition and cleavage by I-*Cre*I, and by other similar eucaryotic mobile intron endonucleases.

## MATERIALS AND METHODS

**Bacterial strains and media:** The following *E. coli* K-12 strains used in this study were kindly supplied by Dr. COLIN MANOIL (Department of Genetics, University of Washington): CC136, [*F128 (lacI$^q$)/Δ(lac-pro), ara, nalA, argE(am), thi, rpoB*]; CJ236, [*dut*1 *ung*1 *thi*-1 *rel*A1/pCJ105(*cam*$^r$ F′)]; CSH116, [*ara mutD5 zae-502::Tn10 Δ(gpt-lac)5 rpsL*]; DB6438, [*lac Δ(lac, pro) XIII, rpoB argE metB mutT198*]; MC1000, [*araD139 Δ(ara, leu)7697 Δ(lac)X74 galE galK thi rpsL*]. Standard growth media were used (MILLER 1992). Where required, they were supplemented with kanamycin (50 μg/ml), carbenicillin (100 μg/ml), chloramphenicol (30 μg/ml), or arabinose (0.2% w/v).

**Plasmid constructions:** pL-S, containing a I-*Cre*I homing site and a kanamycin-resistance gene from Tn*5*, was constructed from pLrec, a 12-kb plasmid containing two *E. coli lacZ* genes that were genetically inactivated by *Xho*I linker insertion or by linker insertion and truncation (HERZING and MEYN 1993). First, a 35-bp oligonucleotide containing a wild-type 24-bp I-*Cre*I homing site sequence and an adjacent *Xba*I site was ligated into pLrec that had been partially digested with *Xho*I. An *Xho*I-*BstB*I fragment containing the Tn*5* kanamycin resistance gene from plasmid pLS200 (SELIGMAN and MANOIL 1994) was then ligated into the *BstB*I and remaining

*Xho*I sites of pLrec*Cre2* to generate pL-S. The orientations of the oligonucleotide and restriction fragment were verified by restriction digestion and DNA sequencing. pLac+ was isolated after pL-S transformation into *E. coli*, followed by screening for the generation of Lac+ colonies on indicator plates. pLac+ appears to have arisen as the result of homologous recombination in the ~600-bp region of homology between the genetically inactive pL-S *lacZ* direct repeats to generate an intact *lacZ* open reading frame.

pB-E and pA-E are compatible plasmids that encode I-*Cre*I endonuclease under control of the promoter and regulatory gene (*araC*) of the *ara*BAD (arabinose) operon of *E. coli*. pB-E was constructed by ligating an *Xba*I-*EcoR*V fragment from pI-*Cre*I, a pET11-based expression vector containing the I-*Cre*I open reading frame (THOMPSON *et al.* 1992; kindly supplied by Dr. DAVID HERRIN, University of Texas-Austin) into *Xba*I- and *Hin*dIII-cleaved pBAD18 DNA (GUZMAN *et al.* 1995). Before ligation, the *Hin*dIII site of pBAD18 was treated with *E. coli* DNA polymerase I Klenow fragment to render it blunt-ended (SAMBROOK *et al.* 1989). pA-E was constructed by transferring a *Cla*I-*Xho*I fragment containing the *ara*BAD regulatory sequences and the I-*Cre*I open reading frame from pB-E to *Cla*I and *Sal*I-cleaved pACYC184 DNA (CHANG and COHEN 1978; New England Biolabs, Beverly, MA).

The I-*Cre*I open reading frame we used differed at seven nucleotide positions from the published sequence (ROCHAIX *et al.* 1985). The revised I-*Cre*I open reading frame sequence, determined by sequencing two independently cloned *C. reinhardtii* chloroplast 23S rRNA restriction fragments, has sequence changes at codons 42 (ACG, *not* GCG) and codons 110–111 (GAACAA, *not* TGGCGG). The remaining A → G base substitution in codon 56 was found only in pI-*Cre*I. This D56G I-*Cre*I variant, hereafter referred to as wild-type I-*Cre*I endonuclease, was also used for the X-ray crystal structure determination of I-*Cre*I (HEATH *et al.* 1997; STEPHENS *et al.* 1997). A site-directed G56D revertant was generated by overlap extension PCR (VILEJO *et al.* 1995) for comparison assays (see Table 1 and RESULTS).

Three other pB-E derivatives were constructed to study potential active site residues. These I-*Cre*I variants, with D20N, S22A or R70A substitutions, were constructed by oligonucleotide-directed mutagenesis. Briefly, the I-*Cre*I open reading frame was transferred to pBluescript II KS(+) (Stratagene, La Jolla, CA) as an *Xho*I-*Xba*I restriction fragment to generate pBS*Cre*. Uracil-containing single-stranded phagemid DNA, isolated by transforming pBS*Cre* into *E. coli* strain CJ236, was then used for oligonucleotide-directed mutagenesis as previously described (KUNKEL *et al.* 1987). The resulting mutations were transferred to pB-E on *Xho*I-*Xba*I restriction fragments and then sequence-verified.

pKS155, a pBluescriptII KS(+)-derived plasmid that contains a wild-type I-*Cre*I homing site was constructed by ligating a 35-bp oligonucleotide containing I-*Cre*I and *Xba*I sites into *Sal*I-cleaved pBluescriptII KS(+) plasmid DNA. Plasmids containing mutant wild-type (pKS162) or mutant (pKS163–169 and pKS172) I-*Cre*I homing sites were constructed by transferring pL-S *Bst*BI-*Eco*RV restriction fragments containing homing sites to *Cla*I- and *Eco*RV-cleaved pBluescriptII KS(+) DNA. Two mutations contained in pKS166 were separated by cleaving with *Asp*700 and I-*Cre*I endonuclease, followed by the recloning of homing site fragments into *Asp*700- and I-*Cre*I-cleaved pKS162 plasmid DNA to generate pKS170 and pKS171. Standard recombinant DNA procedures were used for plasmid constructions and for the isolation and restriction analysis of plasmid DNAs (HOLMES and QUIGLEY 1981; SAMBROOK *et al.* 1989).

**Bacterial manipulations:** Reciprocal recombination was used to transfer the kanamycin resistance gene and adjacent I-*Cre*I homing site of pL-S to the F′ factor F128 (HOLLOWAY and LOW 1996) of *E. coli* strain CC136. The resulting F′S recombinants were selected by isolating kanamycin-resistant colonies after mating, followed by screening for carbenicillin sensitivity to eliminate cointegrants (MILLER 1992). Mutant homing sites were transferred by reciprocal recombination from F′S onto pLac+ by transforming MC1000FS with pLac+. Plasmid DNA from transformants was isolated and transformed into MC1000 to isolate kanamycin-resistant colonies.

**Mutant isolation:** Homing site mutants were obtained by transforming strain CSH116F′S with pB-E plasmid DNA. Carbenicillin- and kanamycin-resistant colonies were purified by restreaking onto antibiotic agar, then cured of plasmid by growth in liquid media containing kanamycin and arabinose (0.2% w/v). Under these growth conditions, wild-type I-*Cre*I-encoding plasmid is efficiently lost. Carbenicillin-sensitive colonies containing putative homing site mutants were then tested for the ability to transfer kanamycin resistance upon mating to MC1000 containing pB-E.

I-*Cre*I endonuclease mutants were generated by passaging pB-E plasmid DNA through mutator strains CSH116 (*mutD5*) or DB6438 (*mutT*). Mutagenized plasmid DNA, isolated from overnight liquid cultures grown from independent colonies, was transformed into MC1000F′S to isolate carbenicillin- and kanamycin-resistant colonies. pB-E plasmid DNA was prepared from colonies purified by restreaking on carbenicillin- and kanamycin-containing agar, then transformed into MC1000F′S. Plasmid DNAs giving comparable transformation efficiencies on carbenicillin- and on carbenicillin- and kanamycin-containing agar were characterized as putative endonuclease mutants.

**Mutant characterization:** The *in vitro* cleavage properties of the mutant I-*Cre*I homing sites were determined as follows. Two hundred nanograms of an *Ahd*I-linearized pKS163–172 plasmid DNA was mixed with 200 ng of *Ahd*I-linearized pKS155 DNA (a 3-kb plasmid containing a wild-type I-*Cre*I homing site) in 25 or 50 μl of 20 mM Tris pH 9.0, 10 mM $MgCl_2$. Digestions were performed with a 2.9-fold molar excess of purified I-*Cre*I for 6 hr at 37°, or with limiting (equimolar) amounts of I-*Cre*I for 20 min at 37°. Cleavage reactions were stopped by placing digestions on ice, followed by the addition of an equal volume of stop solution containing bromphenol blue, 20 mM EDTA and, unless otherwise noted, SDS (to 0.5% w/v) before electrophoresis through 1.2% agarose gels run in 1× TBE buffer (SAMBROOK *et al.* 1989). The gel buffer did not contain SDS. The DNA sequence of mutant I-*Cre*I homing sites was determined by chain terminator sequencing of the homing site region of independent pLac+ recombinants, using an end-labeled 20-base primer as previously described (MONNAT *et al.* 1992).

The overexpression toxicity and/or activity of I-*Cre*I endonuclease mutants was determined by retransforming MC1000F′S with individual pB-E mutant plasmids, followed by selection on agar containing carbenicillin and 0.2% (w/v) arabinose. pB-E plasmids that contain a wild-type I-*Cre*I open reading frame do not generate colonies after transformation under these growth conditions. To determine whether toxicity reflected the retention of partial endonuclease activity, endonuclease mutants were transformed and grown in the presence of 0.2% arabinose then assayed for retention of the F′S kanamycin resistance marker. This was done by restreaking carbenicillin-resistant colonies onto agar containing carbenicillin and kanamycin.

The genetic dominance of I-*Cre*I endonuclease mutants was assayed by determining the ability of mutants to interfere with wild-type I-*Cre*I elimination of the F′S-linked kanamycin resistance marker. MC1000F′S containing a pB-E mutant was transformed with pA-E, a compatible plasmid that expresses wild-type I-*Cre*I endonuclease. Transformants were then selected on agar containing chloramphenicol and carbenicillin

in the presence or absence of kanamycin. Endonuclease mutants with chloramphenicol and kanamycin-resistant to chloramphenicol-resistant (*i.e.*, total) colony ratios of $\leq 10^{-3}$, the background in this system, were scored as recessive.

The DNA sequence of both strands of the I-*Cre*I open reading frame of pB-E mutants was determined by fluorescinated chain terminator sequencing using a set of four, 16–20-base primers that hybridized in or adjacent to the I-*Cre*I open reading frame. These analyses were performed by the Murdock Center (University of Montana, Missoula), by the Department of Biochemistry Sequencing Facility at the University of Washington (Seattle), or by the DNA Sequencing Facility at Rancho Santa Ana Botanic Garden (Claremont, CA).

## RESULTS

**Strategy for mutant isolation:** We have developed and used a genetic assay for I-*Cre*I activity in *E. coli* to determine the sequence requirements for I-*Cre*I homing site recognition and cleavage. This assay system consists of a host strain containing the I-*Cre*I homing endonuclease and homing site, though lacking a copy of the 888-bp I-*Cre*I mobile intron (Figure 1B). Plasmid-borne I-*Cre*I endonuclease is expressed under control of an arabinose-inducible promoter (pB-E; see MATERIALS AND METHODS), while the I-*Cre*I homing site is adjacent to a kanamycin-resistance gene on an F′ factor (F′S). In this system constitutive, low level expression of I-*Cre*I endonuclease in the absence of arabinose induction (GUZMAN *et al.* 1995) leads to the efficient loss of the F′S kanamycin-resistance marker.

**I-*Cre*I homing site mutants:** I-*Cre*I-resistant homing site mutants were isolated by transforming *mutD5* host cells harboring F′S with the I-*Cre*I plasmid pB-E. Kanamycin- and carbenicillin-resistant colonies containing putative mutant homing sites arose at a frequency of $\sim 4 \times 10^{-3}$. To demonstrate that cells carried mutations linked to the F′, mutants were cured of pB-E plasmid, then assayed for the ability to transfer the F′ kanamycin-resistance marker into host cells carrying wild-type pB-E plasmid. Twenty-one mutant I-*Cre*I homing sites identified in this manner were transferred onto pLac+ for further analysis.

DNA sequence analysis of the mutant homing sites revealed at least one mutation in each 24-bp homing site: 20 mutants differed at a single basepair from the wild-type I-*Cre*I homing site sequence, while the remaining mutant contained two base substitutions within the I-*Cre*I homing site. Mutations included base substitutions at positions 5, 8, 14, 15 and 17; and single base deletions at positions 5, 6 and 8. Multiple, independent mutations of a given type were isolated at positions 5, 8 and 15, the same three positions where different molecular types of mutations (base pair deletions and/or base substitutions) were identified (see Figure 2). Two mutations contained in one mutant were separated by recloning and then analyzed individually.

The *in vitro* cleavage sensitivity of mutant I-*Cre*I homing sites was determined by digesting mutant homing site plasmids with excess or limiting amounts of purified I-*Cre*I endonuclease. An equimolar amount of wild-type I-*Cre*I homing site DNA was included in each assay to provide an internal standard for assessing the relative cleavage competence of mutant homing sites under different digest conditions (see MATERIALS AND METHODS). Site-specific cleavage of the I-*Cre*I homing site of the 3.9-kb pBlueScriptII-derived "test" plasmids containing a mutant homing site yielded fragments of 2.4 and 1.5 kb (Figure 3, "test products"). Cleavage of the I-*Cre*I homing site of the 3-kb control plasmid yielded fragments of 1.7 and 1.3 kb (Figure 3, "wt products"). All of the control plasmid and, surprisingly, nine of the 10 homing site mutants were cleaved by excess I-*Cre*I (Figure 3A). The remaining homing site mutant, T8Δ, was resistant to I-*Cre*I cleavage (Figure 3A, lane 5). Digestions performed using just enough I-*Cre*I endonuclease to cleave a majority of wild-type plasmid identified five additional mutant homing sites in addition to T8Δ were cleaved less efficiently than wild-type homing site DNA (Figure 3B; mutants C6Δ, T8C, G14T, A15C, and A15G in lanes 4, 5, 7–10). The remaining four homing site mutants were cleaved as efficiently as wild-type homing site DNA (Figure 3B; A5Δ, A5G, A17T and the double mutant in lanes 3, 6, 11 and 12). Gel shift assays using I-*Cre*I endonuclease have shown that these mutant homing sites are bound less efficiently than is wild-type homing site DNA (K. M. STEPHENS, unpublished results).

During the course of these experiments we found that I-*Cre*I endonuclease requires the addition of SDS to release cleaved homing site DNA. We thus compared the ability of I-*Cre*I endonuclease to cleave and release mutant, as opposed to wild type, homing site DNAs in the *absence* of SDS (Figure 4). Three mutant homing sites displayed SDS-independent cleavage product release (Figure 4; C6Δ, G14T and T8C+A15G, compare

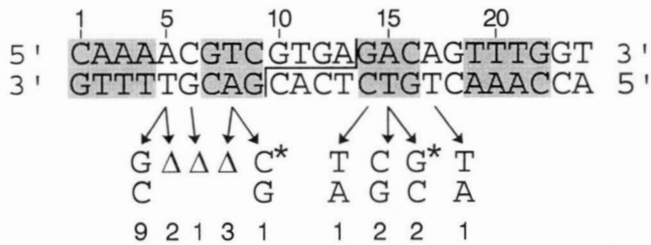


FIGURE 2.—I-*Cre*I homing site sequence and mutations. The sequence of the I-*Cre*I homing site from the *C. reinhardtii* chloroplast 23S rRNA gene is shown, with the asymmetrical cleavage position indicated by the staggered solid line (⌐¬). Regions of partial twofold sequence symmetry are shaded. The positions and types of homing site mutations are indicated below the sequence, with single base deletions indicated by Δ. The two base substitutions found in one mutant homing site are indicated by *. One of the mutations found in the double mutant, the A15G base substitution, was also isolated as a single base substitution. Numbers below each mutation indicate the number of times each mutation was independently isolated.

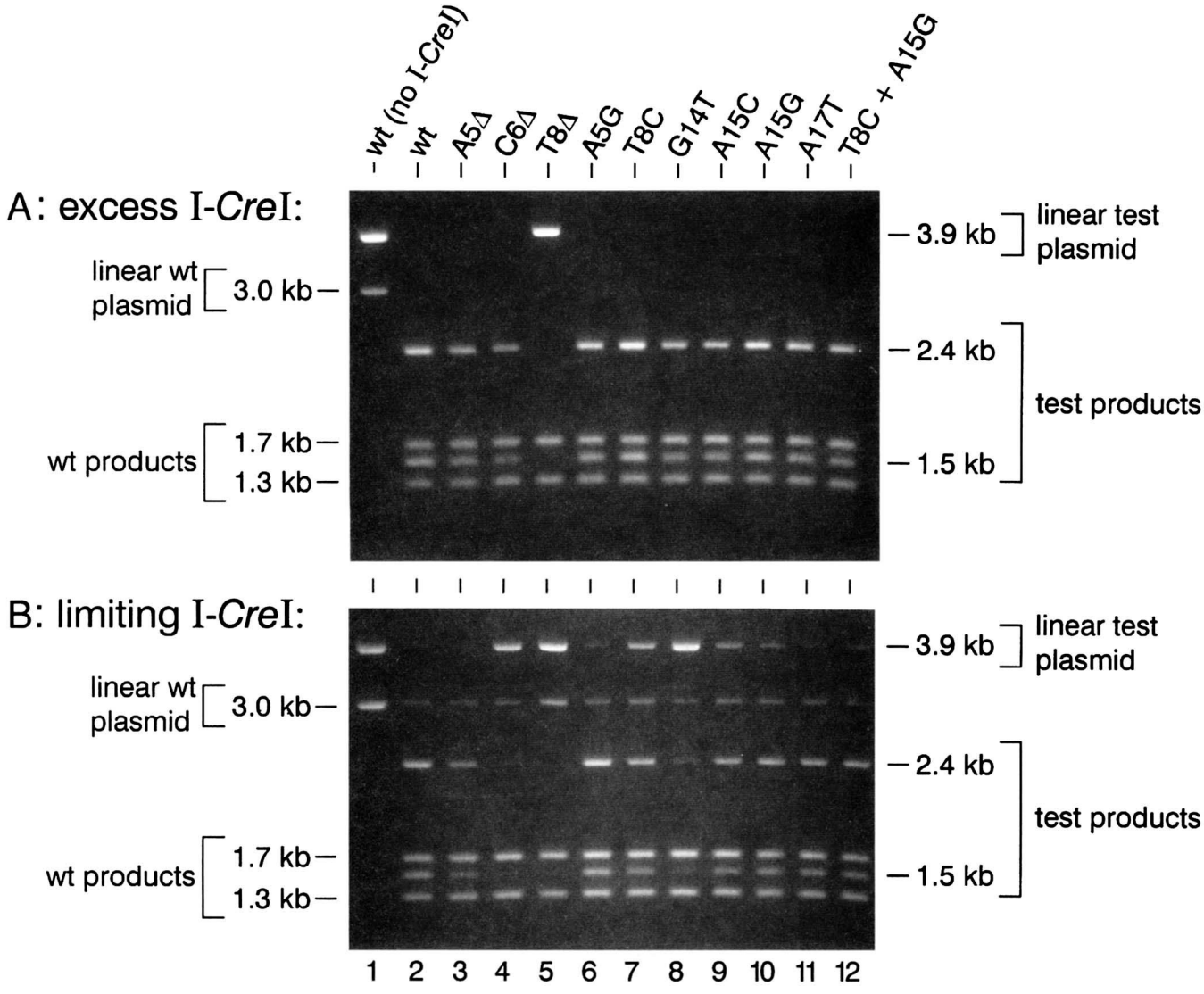


FIGURE 3.—Comparative *in vitro* cleavage sensitivities of mutant I-*Cre*I homing site DNAs. The *in vitro* cleavage properties of mutant I-*Cre*I homing sites contained in plasmids pKS163-172 were determined by digesting mutant homing site DNAs in the presence of an equimolar amount of a wild-type homing site plasmid pKS155 with either excess (A) or equimolar (B) amounts of purified I-*Cre*I endonuclease, before the addition of SDS to 0.5% (v/v) and gel electrophoresis. All of the mutant homing sites with the exception of T8Δ (lane 5) were cleaved to completion in the presence of excess I-*Cre*I endonuclease (A). Five homing site mutants, C6Δ (lane 4), T8C (lane 7), G14T (lane 8), A15C (lane 9), and A15G (lane 10), in addition to T8Δ, were cleaved less efficiently than wild-type homing site DNA in digestions performed using limiting amounts of I-*Cre*I (B).

"+SDS" and "−SDS" lanes). Each of the mutations contained in the double mutant homing site, T8C+A15G, displayed partial SDS-independent cleavage product release (Figure 4, compare lanes 7 and 8 with lanes 9–12).

**I-*Cre*I endonuclease mutants:** We also used the *E. coli*-based selection system described above to isolate and characterize I-*Cre*I endonuclease mutants. Kanamycin- and carbenicillin-resistant colonies containing potential endonuclease mutants arose at a frequency of $\sim 8 \times 10^{-4}$. Twenty-nine plasmid DNAs that generated kanamycin- and carbenicillin-resistant colonies at high frequency upon retransformation were further characterized. The *in vivo* toxicity, activity (Figure 5A) or dominance (Figure 5B) of each mutant was determined in *E. coli*, followed by DNA sequencing of the I-*Cre*I open reading frame. The 29 endonuclease mutants, including 26 isolated in *E. coli* and three generated by site-directed mutagenesis, could be placed into six phenotypic classes (Table 1, Figure 6).

The *in vivo* toxicity and activity assays of I-*Cre*I endonuclease mutants took advantage of the arabinose-inducible promoter regulating the I-*Cre*I open reading frame of pB-E (Figure 5A). Plasmid containing a wild-type I-*Cre*I endonuclease gene failed to generate carbenicillin-resistant colonies after transformation and growth on agar containing arabinose, as did a G56D variant. All of the I-*Cre*I endonuclease mutants, in con-

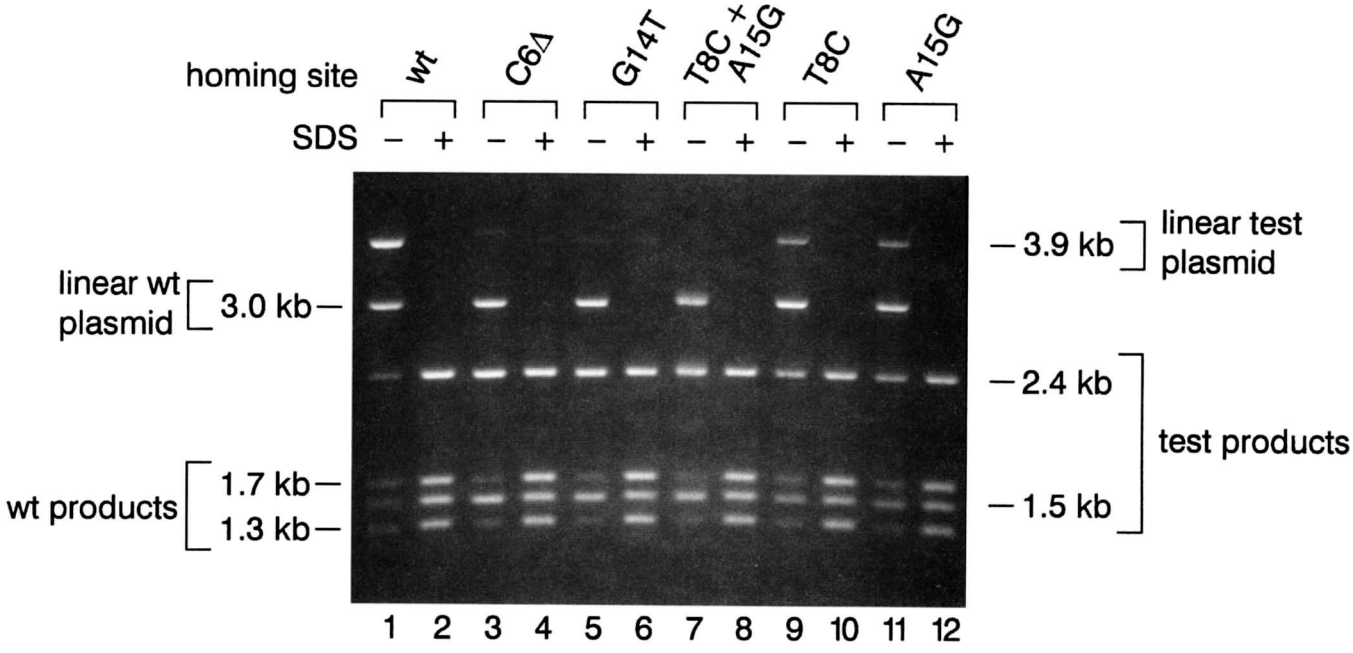


FIGURE 4.—SDS-independent release of cleaved I-*Cre*I homing site DNA. Wild-type I-*Cre*I homing site DNA is cleaved, though does not rapidly release cleavage products, in the absence of SDS (compare lanes 1 and 2). The release of cleaved mutant homing site DNA in the absence of added SDS is shown for four different homing site mutants: a base deletion (C6Δ, lanes 3 and 4), a base substitution (G14T, lanes 5 and 6), a double mutant homing site (T8C+A15G, lanes 7 and 8) and the base substitutions contained in this site (T8C, lanes 9 and 10; and A15G, lanes 11 and 12).

trast, produced carbenicillin-resistant colonies on agar containing arabinose: seven of the 29 mutants displayed normal colony size and morphology (Table 1, "overexpression nontoxic" mutant classes I and II), while the remaining 22 formed small, irregular colonies (Table 1, "overexpression toxic" mutant classes III–V). We determined whether overexpression toxic I-*Cre*I endonuclease mutants retained partial endonuclease activity by assaying colonies grown in the presence of arabinose for retention of the F'S kanamycin resistance marker.

Three of the six mutants that were isolated in *E. coli* and were neither overexpression toxic or active (Table 1, class I) contained mutations leading to premature translation termination (E80 stop and L88 stop mutants), or a large deletion that eliminated most of the I-*Cre*I open reading frame. Three additional mutants in this class were not characterized at the DNA sequence level. The remaining nontoxic mutant (L10 stop; Table 1, mutant class II) contained a translation termination codon at I-*Cre*I residue 10. This mutant retained residual activity and was presumably active by virtue of stop codon read-through or suppression.

Twenty-two different mutations were toxic when overexpressed (Table 1, mutant classes III–VI). These included 18 different base substitutions, two 6-bp insertions and two frameshifts (Table 1 and Figure 6). One base substitution generating a D75G substitution was independently isolated three times, and one insertion generating a $+LK_{99\text{-}100}$ duplication was independently isolated twice. Toxic mutants that *retained* endonuclease activity (mutant classes IV–VI), including two frameshift mutations that led to premature translation termination, were located throughout the I-*Cre*I protein. In contrast, mutations that *inactivated* endonuclease activity were located in the N-terminal two-thirds of the I-*Cre*I protein. One of these, at residue 47, was in a glutamine codon that is highly conserved among related mobile intron-encoded proteins (Figure 6).

The genetic dominance of I-*Cre*I mutants was determined by assaying their ability to block wild-type I-*Cre*I-dependent elimination of the F'S kanamycin-resistance marker (Figure 5B). I-*Cre*I mutants were classified as "recessive" if they displayed a ratio of kanamycin-resistant colonies to total colonies (in the absence of arabinose) of $\leq 10^{-3}$, the background ratio in this assay. Four of the 13 I-*Cre*I mutants displayed overexpression-associated toxicity, activity and kanamycin resistant:total ratios $\geq$100-fold above background, and were thus classified as "dominant" (Table 1, mutant class V). These mutants had single amino acid substitutions at I-*Cre*I residues 34, 35, 69 and 142 (Figure 6).

Only one potential active site mutant, Q47H, was isolated by genetic selection in *E. coli*. Three other potential active site mutants were constructed by site-directed mutagenesis and analyzed in *E. coli*. D20N and R70A substitution mutants were toxic, though neither active or dominant. These mutants thus resemble Q47H and other class III mutants (Table 1). An S22A mutant was overexpression toxic and active in the absence of arabinose induction. This mutant was classified separately

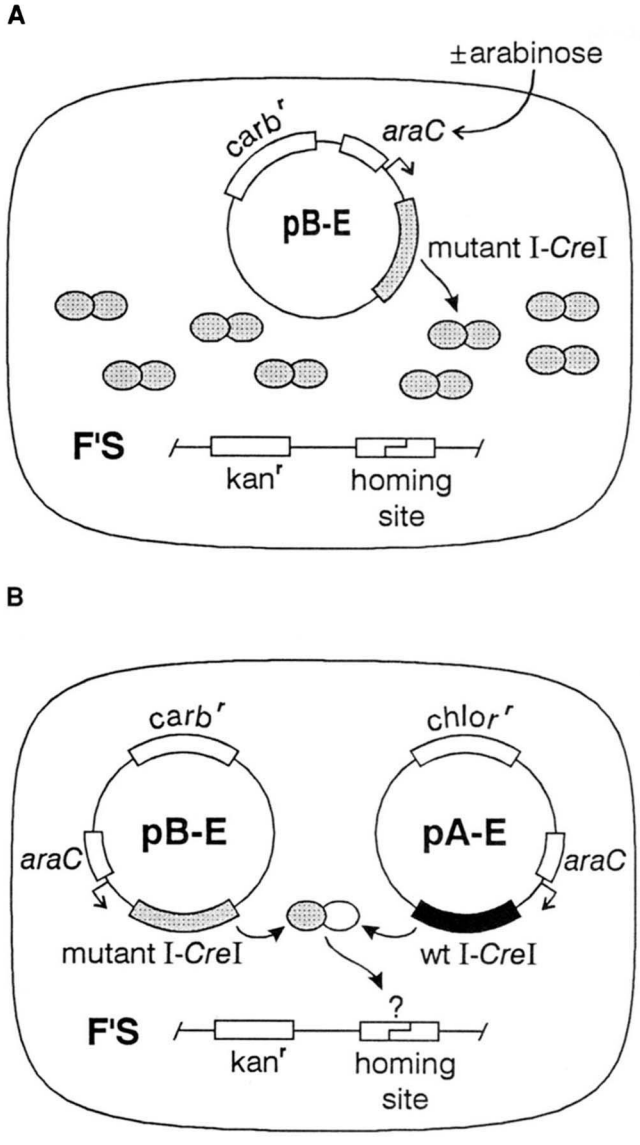


FIGURE 5.—Assays for overexpression toxicity, activity or dominance of I-*Cre*I endonuclease mutants. (A) The toxicity and/or activity of putative I-*Cre*I endonuclease mutants was determined by overexpressing mutant endonucleases in the presence of an F′S I-*Cre*I homing site, followed by assays for growth, colony size, and morphology and retention of the kanamycin resistance marker. (B) Genetic dominance was assayed by determining the ability of I-*Cre*I endonuclease mutants to interfere with the ability of coexpressed wild-type I-*Cre*I to cleave and eliminate the F′S kanamycin resistance marker. Mutant and wild-type I-*Cre*I endonucleases were expressed from compatible plasmids (pB-E and pA-E, respectively). I-*Cre*I mutants were classified as recessive, if they displayed a kanamycin resistance-to-sensitivity ratio of $\leq 10^{-3}$, the background ratio in this assay.

**TABLE 1**

**I-*Cre*I endonuclease mutants**

| Mutant class | Sequence[a] | Toxicity[b] | Activity[c] | Dominance[d] |
|---|---|---|---|---|
| — | D56[e]<br>G56 | ++ | ++ | NA |
| I | E80 stop<br>L88 stop<br>Δ ~1 kb[f]<br>Three additional[g] | − | − | − |
| II | L10 stop | − | + | − |
| III | D20N[e]<br>L39R<br>Q47H<br>R51G<br>R70A[e]<br>D75G[h] (3)<br>L91R<br>K98Q<br>$+K_{102}Q_{103}$ | + | − | − |
| IV | K28T<br>Y76C<br>FS 88<br>FS 101<br>$+L_{99}K_{100}$ (2)<br>W128G<br>V129G<br>D137E | + | + | − |
| V | K34N<br>F35C<br>D69G<br>K142Q | + | + | + |
| VI | S22A[e] | + | ++ | NA |

[a] Numbers in parentheses indicate independent occurrences; + indicates two amino acid insertions after amino acids 101 and 98, respectively; FS indicates single base pair insertions at the indicated codons. Each frame-shift mutation generates a 109 residue protein with nine or 21 novel amino acids, respectively. All mutants contain the G56 substitution.

[b] ++, +, − indicate a lack of colonies, small irregularly shaped colonies, and normal sized colonies, respectively, on arabinose-containing media.

[c] ++, +, − indicate active without arabinose present, active after exposure to arabinose, and inactive after exposure to arabinose, respectively.

[d] NA, not applicable.

[e] Obtained by oligonucleotide-directed mutagenesis.

[f] Restriction analysis indicated >89% of coding region deleted.

[g] No molecular analysis performed on these mutants.

[h] Only mutant demonstrating variable response in activity assay: >88% of colonies grown in the presence of arabinose displayed kanamycin resistance.

(Table 1, class VI) as its *in vivo* behavior resembled none of the other endonuclease mutants we had isolated or constructed.

## DISCUSSION

**Genetic analysis of a eukaryotic homing system:** We have used a genetic selection system in *E. coli* to study sequence requirements for the initial step of intron homing by the I-*Cre*I mobile intron homing system. This system contained the I-*Cre*I endonuclease and its homing site DNA, though it lacked the 888-bp mobile intron (ROCHAIX *et al.* 1985; HERRIN *et al.* 1990; DÜRRENBERGER

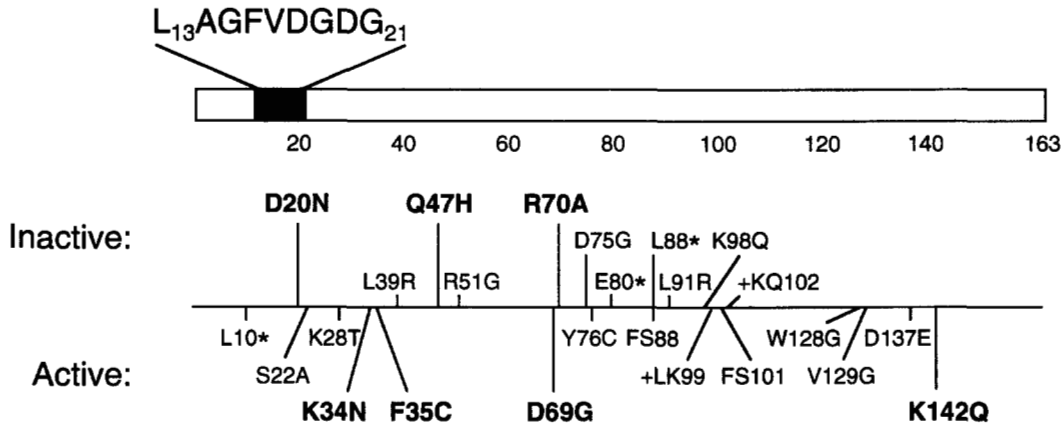


FIGURE 6.—Position and activity of mutations in the I-*Cre*I endonuclease open reading frame. The open box (top) depicts the 163 amino acid residue I-*Cre*I endonuclease. The position and amino acid sequence of the I-*Cre*I LAGLIDADG motif at residues 13–21 are shown by the filled box and amino acid sequence (left end, top). The location and activity of 25 I-*Cre*I open reading frame mutations is shown beneath the depicted protein: mutations that inactivate I-*Cre*I endonuclease function are shown above the horizontal line, while those that retain partial activity are shown below the line. The position of each mutation in the I-*Cre*I open reading frame is given by the number indicating the affected codon(s). Letters indicate the amino acid change(s) in each mutant at the affected codon(s). Mutations that generate a translation stop codon are indicated by *. The insertion of amino acid residues is indicated by a +, followed by the residues inserted and the position of the I-*Cre*I codon N-terminal to the inserted amino acid residues. The positions of two frameshift mutations are indicated by FS followed by a number indicating the altered codon. Two site-directed mutants that inactive endonuclease activity are shown above the line in bold (D20N and R70A), along with the inactivating Q47H mutant that was isolated in *E. coli*. An S22A site-directed mutant that retained activity is shown below the line, as are four I-*Cre*I endonuclease mutants, in large bold type, that retained partial activity and displayed genetic dominance.

and ROCHAIX 1991, 1993; THOMPSON and HERRIN 1991; THOMPSON *et al.* 1992). Our aim was to identify I-*Cre*I homing site basepairs and I-*Cre*I endonuclease residues that play mechanistically important roles in homing site recognition and cleavage.

**I-*Cre*I homing site sequence requirements:** Twenty-one independent mutants were isolated that contained six single base substitutions and three single base deletions at six different positions in the I-*Cre*I homing site sequence. Five of the nine different homing site mutations were isolated as independent mutants on two or more occasions. Twelve of the 21 homing site mutations were transitions, a slightly smaller fraction than expected for the *mutD5* mutator in rich medium (SCHAAPER 1988; MILLER 1992).

Molecular modeling of I-*Cre*I homing site DNA bound to the I-*Cre*I endonuclease indicates the potential for DNA-protein contacts extending over at least 20 bp of the I-*Cre*I homing site sequence (HEATH *et al.* 1997). It is interesting that we isolated mutations only at homing site positions 5–8 and 14–17, and that these mutations displayed a range of sensitivities to *in vitro* cleavage by purified I-*Cre*I endonuclease. Single base deletions, in general, more strongly inhibited cleavage than did single base substitutions. The homing site mutations did not consistently target basepairs that would alter the limited twofold sequence symmetry of the I-*Cre*I homing site (Figure 2). These results indicate that basepairs 5–8 and 14–17 of the I-*Cre*I homing site are proportionately more important role in site recognition and cleavage than are flanking basepairs. We isolated no mutations in the central, asymmetrically cleaved 4-bp region between homing site basepairs 5–8 and 14–17. This result was anticipated, as molecular modeling did not suggest that critical DNA-protein contacts would occur in this 4-bp region. Moreover, we have shown that site-directed mutagenesis of two of these four basepairs (at positions 12 and 13, to create symmetric four base overhangs; see Figure 2) does not affect homing site binding or cleavage (HEATH *et al.* 1997, and K. M. STEPHENS, unpublished results).

One surprising finding was the ability of different I-*Cre*I homing site mutations to relax the requirement for adding SDS to digestions to liberate the cleaved homing site products. I-*Cre*I endonuclease forms a tight, though noncovalent, enzyme-product complex with cleaved native homing site DNA. This explains the need to add at least an equimolar amount of I-*Cre*I endonuclease to achieve complete digestion of native homing site DNA. Many mutant homing sites can, however, be digested to completion with smaller amounts of purified I-*Cre*I endonuclease (K. M. STEPHENS, unpublished results). This suggests that the stability of the postcleavage complex can be reduced by specific homing site mutations, and that these mutations identify potential DNA-protein contacts that are important in generating or stabilizing the enzyme-product complex. A similar, though less dramatic, association of endonuclease and cleaved homing site DNA has been noted for other homing endonucleases (*e.g.*, I-*Sce*I), though it is not known in these instances if homing site mutations alter the rate of product release (MONTEILHET *et al.* 1990;

PLESSIS *et al.* 1992; PERRIN *et al.* 1993). One postulated role for these endonuclease-product complexes is to facilitate intron insertion by protecting cleaved homing site DNA ends from exonucleolytic degradation. It is not known, however, whether I-*Cre*I or I-*Sce*I form stable, noncovalent complexes with their cleaved homing site DNAs *in vivo*.

Our mutant homing site data indicate that selection in *E. coli* provides a useful way to isolate homing site mutations that reduce or abolish homing site cleavage. This selection system also allowed us to fortuitously recover homing site mutants that altered the rate of cleavage product release. The partial sensitivity of many of our homing site mutants to I-*Cre*I digestion *in vitro* suggests that our selection system is able to reveal small decrements in the ability of I-*Cre*I to cleave F'S I-*Cre*I homing site DNA. Comparable mutant homing sites for a second mobile intron endonuclease, I-*Dmo*I, have been selected *in vitro* on the basis of resistance to cleavage with limiting amounts of endonuclease (AAGAARD *et al.* 1997). Further biochemical analyses of mutant I-*Cre*I homing sites, in conjunction with a high-resolution co-crystal structure of I-*Cre*I bound to homing site DNA, should allow us to identify the specific DNA-protein contacts involved in homing site recognition, in cleavage and in the generation of a tight, though noncovalent, enzyme-product complex.

**Functional organization of the I-*Cre*I endonuclease:** We used a modification of the genetic selection system described above to isolate and characterize I-*Cre*I endonuclease mutants. A combination of toxicity, activity and genetic dominance assays in *E. coli* were used to further characterize and identify the most potentially informative I-*Cre*I endonuclease mutants (*cf.* PAKULA and SAUER 1989). Six different mutant classes could be identified among 29 endonuclease mutants. The mutants isolated in *E. coli* contained 26 unique mutations in the I-*Cre*I open reading frame (Table 1 and Figure 6) that were close to the expected spectra for the mutator strains that we employed: 11 of 14 mutations isolated by passaging through *mut T* were A → C transversions, whereas half (five of 10) of the *mutD*-derived mutants consisted of A → G transitions (SCHAAPER 1988; MILLER 1992).

Mutations that inactivated endonuclease function included six single amino acid substitutions that were largely nonconservative, a duplication of the lysyl-glutamine dipeptide at positions 102–103, and two chain termination codons at amino acid residues 80 and 88. These mutations are clustered in the N-terminal two-thirds of the I-*Cre*I sequence, the portion of the open reading frame that is retained in two independent, enzymatically active frameshift mutants. These results indicate that homing site recognition and cleavage are mediated by the N-terminal two-thirds of I-*Cre*I, as suggested by the X-ray crystal structure of native I-*Cre*I (Figure 7; see also below). Many of the amino acid substitutions that inactivate I-*Cre*I endonuclease function retain toxicity upon overexpression in *E. coli*. This suggests that toxicity and endonuclease activity are not tightly linked. I-*Cre*I endonuclease has a strong net basic charge, with a pI of ~10.5. Thus enzymatically inactive I-*Cre*I might retain the potential to act as a DNA binding protein that could alter chromosome structure or gene expression in *E. coli*. A complementary set of results have been obtained for mutants of I-*Ceu*I, where mutants originally selected on the basis of reduced toxicity displayed a range of enzyme activities (TURMEL *et al.* 1997).

A particularly interesting subset of I-*Cre*I mutants displayed overexpression activity and toxicity, and were able to interfere with the ability of wild-type I-*Cre*I to cleave homing site DNA in *E. coli*. Each of these dominant mutants had nonconservative amino acid substitutions in the I-*Cre*I open reading frame. The "overexpression active" phenotype of these four mutants suggests that each has a lower affinity for homing site DNA than wild-type I-*Cre*I, and that dominance results from the ability of mutants to sequester wild-type I-*Cre*I endonuclease in heterodimers that can no longer efficiently recognize or cleave homing site DNA. The positions of mutations in three of the four dominant mutants support this argument. The K34N and F35C substitutions occur in a loop between sheets $\beta 1$ and $\beta 2$ that has the potential to contact specific base pairs at the ends of the I-*Cre*I homing site. The D69G substitution, in contrast, is positioned near the substrate binding surface and potential active site residue R70. The remaining dominant mutant, K142Q, is enigmatic: it is located in the C-terminus of I-*Cre*I where no additional mutations have been found, and where the native I-*Cre*I crystal structure is disordered (Figure 7 and HEATH *et al.* 1997). Mutant forms of I-*Cre*I that retained the ability to form dimers with wild-type I-*Cre*I may be useful for analyzing the role of subunit communication in homing site binding, cleavage and product release (see, *e.g.*, STAHL *et al.* 1996).

We were surprised that we isolated no inactivating mutations in the single copy of the LAGLIDADG motif at I-*Cre*I residues 13–21. This motif has been found in many other mobile intron endonucleases and maturases, where it has been postulated to play a role in substrate binding and/or cleavage (MICHEL *et al.* 1982; HENSGENS *et al.* 1983; WARING *et al.* 1983; HODGES *et al.* 1992; MUELLER *et al.* 1993; PIETROKOVSKI 1994; BELFORT and PERLMAN 1995; GIMBLE and STEPHENS 1995; HENKE *et al.* 1995; PIETROKOVSKI 1997). The X-ray crystal structure of native I-*Cre*I indicates that the LAGLIDADG motif of I-*Cre*I plays *two* roles: it forms the dimer interface and contributes amino acid residues, most notably D20, to a putative endonuclease active site in each monomer. A comparable dual role that we predicted for mobile intron endonucleases containing two copies of the LAGLIDADG motif (HEATH *et al.* 1997) has been re-

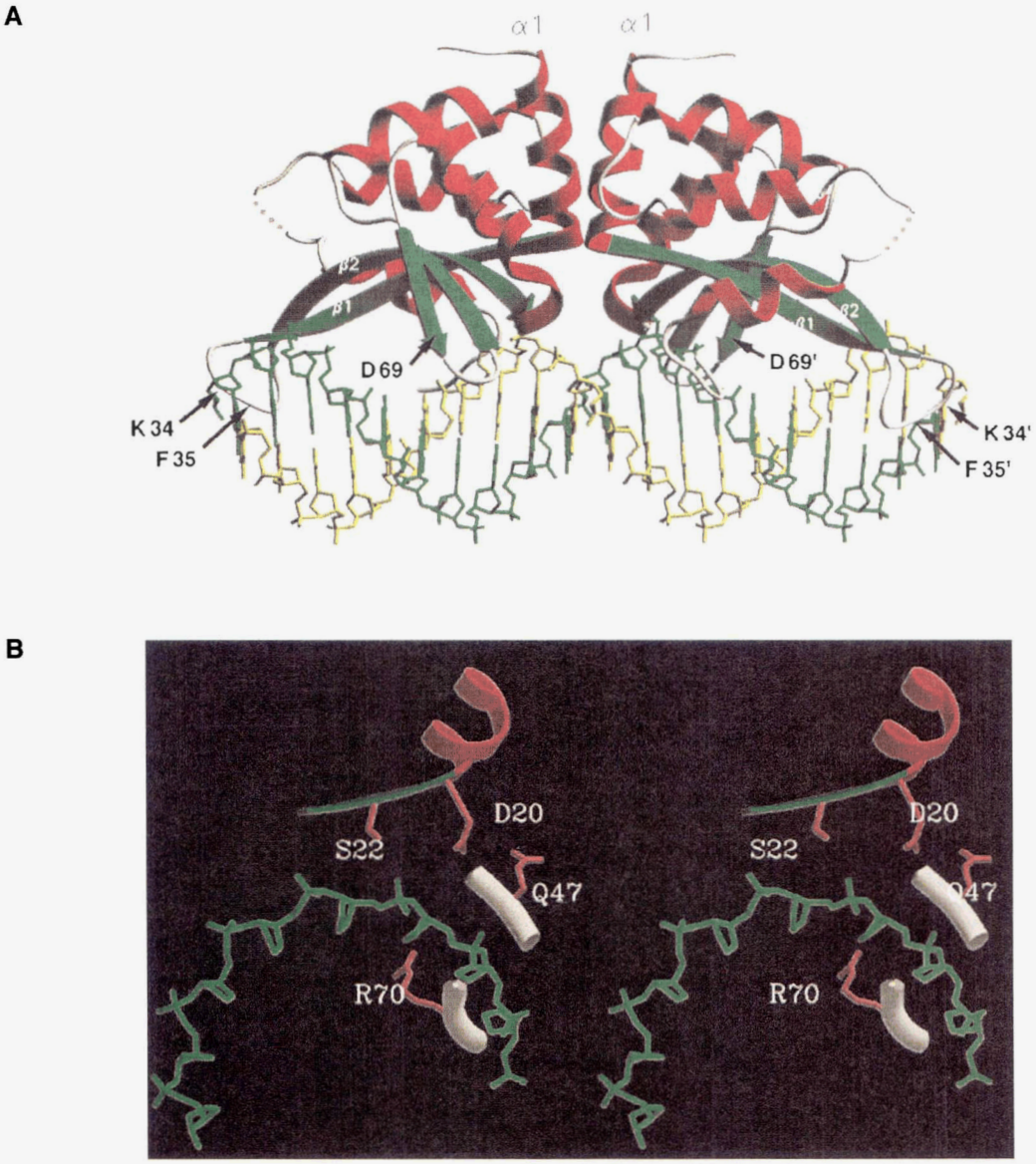


FIGURE 7.—Model of native I-*Cre*I endonuclease docked to I-*Cre*I homing site DNA and active site mutants. (A) The active form of I-*Cre*I is an endonuclease homodimer, in which the four antiparallel $\beta$ strands of each monomer form a DNA binding surface or saddle that is ~20 Å wide and 70 Å long. The LAGLIDADG motifs of the two monomers ($\alpha 1/\alpha 1'$) form a novel $\alpha$-helical dimer interface, and contribute aspartic acid residues D20 and D20′ to a predicted bipartite active site at the center of the homodimer. The I-*Cre*I DNA binding surface makes contact with at least 20 bp of the I-*Cre*I homing site. The most distant potential contacts are by loops between strands $\beta 1$ and $\beta 2$ that can contact homing site bp 2 and 3 and 19–21 via the major groove. The positions of three dominant mutants (K34N, F35C and D69G) are indicated in the ribbon diagram by arrows. (B) Stereo diagram of the postulated I-*Cre*I endonuclease active site residues in an I-*Cre*I monomer. The green strand indicates the position of the phosphodiester backbone of one strand of the homing site DNA. Potential active site residues are indicated in red (D20, S22, Q47 and R70). The phenotypes of D20, Q47 and R70 mutants indicate that these three residues are likely to be active site residues, whereas S22 is not (see text). The phosphodiester bond cleaved is located at the center of the triangle formed by D20, S22 and Q47.

cently confirmed with publication of the structure of PI-*Sce*I, an intein-encoded mobile intron endonuclease (DUAN *et al.* 1997).

The X-ray crystal structure of I-*Cre*I and phylogenetic data on related mobile intron-encoded proteins indicate that four different amino acid residues may contribute to the I-*Cre*I active site: D20 in the LAGLIDADG motif, and residues S22, Q47 and R70 (HEATH *et al.* 1997; TURMEL *et al.* 1997). We isolated one I-*Cre*I endonuclease mutant in *E. coli*, a Q47H substitution that was enzymatically inactive (Table 1), and thus similar to an inactivating substitution found at the comparable residue of I-*Ceu*I (Q93R; TURMEL *et al.* 1997). To determine the functional importance of the other three potential active site residues, we constructed site-directed mutations to generate D20N, S22A and R70A amino

acid substitutions. The phenotypes of the D20N and R70A substitutions were identical to those of Q47H: toxic and inactive, though not dominant when expressed in *E. coli*. These two mutants thus resembled mutants (I-*Ceu*I E66K and K116Q, respectively) that have been shown to inactive I-*Ceu*I endonuclease activity (TURMEL *et al.* 1997). The S22A I-*Cre*I mutant, in contrast, was active in the absence of arabinose induction. However, unlike either wild-type or G56D I-*Cre*I, S22A could generate colonies on agar containing arabinose. These results suggest that S22 is unlikely to play a functionally important role in the I-*Cre*I active site.

The I-*Cre*I active site model we favor has two symmetry-related metal binding sites, formed by D20, Q47 and R70 in each monomer, positioned ~10Å apart at the center of the I-*Cre*I DNA binding surface. These active sites are appropriately positioned to simultaneously cleave homing site DNA bound in the presence of $Mg^{2+}$ across the minor groove to liberate four base, 3′OH-extended cohesive ends (ANDERSON 1993; AGGARWAL 1995; HEATH *et al.* 1997). The postulated I-*Cre*I active site would thus resemble the active sites of several well characterized Type II restriction endonucleases in that it is bipartite, with each metal-binding site formed by acidic (D20), basic (R70) and polar (Q47) residues that are widely distributed in the protein primary sequence (AGGARWAL 1995; HEATH *et al.* 1997). In contrast, the recently determined structure of the PI-*Sce*I mobile intron endonuclease appears to contain a single active site consisting of two aspartic acid residues and a lysine. The single PI-*Sce*I active site may act sequentially to cleave both strands of the PI-*Sce*I homing site to generate four base, 3′OH-extended cohesive ends (DUAN *et al.* 1997).

**Summary:** We have developed and used a simple genetic selection system in *E. coli* to explore the sequence requirements for homing site recognition and cleavage by the I-*Cre*I eukaryotic mobile intron endonuclease. This genetic approach should be applicable to the analysis of other eukaryotic mobile intron homing systems (see, *e.g.*, TURMEL *et al.* 1997) and may have practical applications. For example, genetic selection could be used to isolate mutant forms of I-*Cre*I, or of other mobile intron endonucleases, that are able to recognize and cleave variant or novel homing sites. Such mutant proteins could be used to target or recruit endonucleolytic or other specific biochemical functions to genes containing these target sites in cells. The feasibility of this approach is being explored in plant and mammalian cells, where it is known that at least four mobile intron endonucleases, I-*Cre*I, I-*Sce*I, PI-*Sce*I and I-*Ppo*I, can find and cleave their cognate homing sites with high specificity (reviewed in JASIN 1996; R. J. MONNAT, JR., F. M. H. HACKMANN, and M. A. CANTRELL, unpublished results).

We thank GINGER ARMBRUST, FRANZ DÜRRENBURGER, ELIZABETH HARRIS, PAT HEATH, DAVID HERRIN, JEAN-DAVID ROCHAIX and BARRY STODDARD for helpful discussions and for communicating results before publication. ELIZABETH HARRIS, DAVID HERRIN, COLIN MANOIL, STEVE MEYN, JEAN-DAVID ROCHAIX and BETH TRAXLER kindly provided materials or strains. PATRICK COLLINS performed *E. coli* phenotyping assays on several mutants. ALDEN HACKMANN generated Figures 1–5 and PAT HEATH generated Figure 7. BAEK KIM, COLIN MANOIL and ANDREW TAYLOR provided useful comments on a draft of this manuscript. This work was supported by grants from the National Institutes on Aging, the National Cancer Institute and the University of Washington Royalty Research Fund to R.J.M., Jr. L.M.S. and K.M.S. were supported by National Institutes of Health Training Grants T32 CA09437 and T32 ES07032, respectively. J.H.S. was supported by National Science Foundation REU grant BIR-9531713.

## LITERATURE CITED

AAGAARD, C., M. J. AWAYEZ and R. A. GARRETT, 1997 Profile of the DNA recognition site of the archeal homing endonuclease I-*Dmo*I. Nucleic Acids Res. **25:** 1523–1530.

AGGARWAL, A. K., 1995 Structure and function of restriction endonucleases. Curr. Opin. Struct. Biol. **5:** 11–19.

ANDERSON, J. E., 1993 Restriction endonucleases and modification methylases. Curr. Opin. Struct. Biol. **3:** 24–30.

BELFORT, M., and P. S. PERLMAN, 1995 Mechanisms of intron mobility. J. Biol. Chem. **270:** 30237–30240.

BELFORT, M., and R. ROBERTS, 1997 Homing endonucleases—keeping the house in order. Nucleic Acids Res. **25:** 3379–3388.

BELFORT, M., M. E. REABAN, T. COETZEE and J. Z. DALGAARD, 1995 Prokaryotic introns and inteins: a panoply of form and function. J. Bacteriol. **177:** 3897–3903.

CHANG, A. C. Y., and S. N. COHEN, 1978 Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:** 1141–1156.

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

DÜRRENBERGER, F., and J.-D. ROCHAIX, 1991 Chloroplast ribosomal intron of *Chlamydomonas reinhardtii*: *in vitro* self-splicing, DNA endonuclease activity and *in vivo* mobility. EMBO J. **10:** 3495–3501.

DÜRRENBERGER, F., and J.-D. ROCHAIX, 1993 Characterization of the cleavage site and the recognition sequence of the I-*Cre*I DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. **236:** 409–414.

DÜRRENBERGER, F., A. J. THOMPSON, D. L. HERRIN and J.-D. ROCHAIX, 1996 Double strand break-induced recombination in *Chlamydomonas reinhardtii* chloroplasts. Nucleic Acids Res. **24:** 3323–3331.

GIMBLE, F. S., and B.W. STEPHENS, 1995 Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of PI-*Sce*I endonuclease. J. Biol. Chem. **270:** 5849–5856.

GUZMAN, L., D. BELIN, M. J. CARSON and J. BECKWITH, 1995 Tight regulation, modulation, and high-level expression by vectors containing the arabinose $P_{BAD}$ promoter. J. Bacteriol. **177:** 4121–4131.

HEATH, P. J., K. M. STEPHENS, R. J. MONNAT, JR. and B. L. STODDARD, 1997 The structure of I-*Cre*I: a Group I intron-encoded homing endonuclease. Nature Struct. Biol. **4:** 468–476.

HENKE, R. M., R. A. BUTOW and P. S. PERLMAN, 1995 Maturase and endonuclease functions depend on separate conserved domains of the bifunctional protein encoded by the group I intron aI4α of yeast mitochondrial DNA. EMBO J. **14:** 5094–5099.

HENSGENS, L. A. M., L. BONEN, M. DE HAAN, G. VAN DER HORST and L. A. GRIVELL, 1983 Two intron sequences in yeast mitochondrial COX1 gene: homology among URF-containing introns and strain-dependent variation in flanking exons. Cell **32:** 379–389.

HERRIN, D. L., Y. CHEN and G. W. SCHMIDT, 1990 RNA splicing in *Chlamydomonas* chloroplasts: self-splicing of 23S preRNA. J. Biol. Chem. **265:** 21134–21140.

HERZING, L. B. K., and M. S. MEYN, 1993 Novel *LacZ*-based recombination vectors for mammalian cells. Gene **137:** 163–169.

HODGES, R., F. B. PERLER, C. J. NOREN and W. E. JACK, 1992 Protein splicing removes intervening sequences in an archaea DNA polymerase. Nucleic Acids Res. **20:** 6153–6157.

HOLLOWAY, B., and K. B. LOW, 1996 F-prime and R-prime factors, pp. 2413-2420 in *Escherichia coli and Salmonella: Cellular and Molecular Biology,* Ed. 2, edited by F. C. NEIDHARDT, R. I. CURTISS, J. L. INGRAHAM, E. C. C. LIN, K. B. LOW, JR., *et al.* ASM Press, Washington, D.C.

HOLMES, D. S., and M. QUIGLEY, 1981 A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:** 193–197.

JASIN, M., 1996 Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet **12:** 224–228.

KUNKEL, T. A., J. D. ROBERTS and R. A. ZAKOUR, 1987 Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154:** 367–382.

LAMBOWITZ, A. M., and M. BELFORT, 1993 Introns as mobile genetic elements. Annu. Rev. Biochem. **62:** 587–622.

LEMIEUX, C., J. BOULANGER, C. OTIS and M. TURMEL, 1989 Nucleotide sequence of the chloroplast large subunit rRNA gene from *Chlamydomonas reinhardtii.* Nucleic Acids Res. **17:** 7997.

MICHEL, F., A. JACQUIER and B. DUJON, 1982 Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie **64:** 867–881.

MILLER, J. H., 1992 *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria,* Ed. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MONNAT, R. J., JR., A. F. M. HACKMANN and T. A. CHIAVEROTTI, 1992 Nucleotide sequence analysis of human hypoxanthine phosphoribosyltransferase gene deletions. Genomics **13:** 777–787.

MONTEILHET, C., A. PERRIN, A. THIERRY, L. COLLEAUX and B. DUJON, 1990 Purification and characterization of the *in vitro* activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res. **18:** 1407–1413.

MUELLER, J. E., M. BRYK, N. LOIZOS and M. BELFORT, 1993 Homing endonucleases, pp. 111-143 in *Nucleases,* Ed. 2, edited by S. M. LINN, R. S. LLOYD and R. J. ROBERTS. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

PAKULA, A. A., and R. T. SAUER, 1989 Genetic analysis of protein stability and function. Annu. Rev. Genet. **23:** 289–310.

PERRIN, A., M. BUCKLE and B. DUJON, 1993 Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. EMBO J. **12:** 2939–2947.

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

PIETROKOVSKI, S., 1997 Modular organization of inteins and C-terminal autocatalytic domains. Protein Sci. (in press).

PLESSIS, A., A. PERRIN, J. E. HABER and B. DUJON, 1992 Site-specific recombination determined by I-*Sce*I, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics **130:** 451–460.

ROBERTS, R., and S. HALFORD, 1993 Type II restriction endonucleases, pp.35–88 in *Nucleases,* Ed.2, edited by S. LINN, R. LLOYD and R. J. ROBERTS. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

ROCHAIX, J. D., and P. MALNOE, 1978 Anatomy of the chloroplast ribosomal DNA of *Chlamydomonas reinhardtii.* Cell **15:** 661–670.

ROCHÁIX, J. D., M. RAHIRE and F. MICHEL, 1985 The chloroplast ribosomal intron of *Chlamydomonas reinhardtii* codes for a polypeptide related to mitochondrial maturases. Nucleic Acids Res. **13:** 975–984.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual,* Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SCHAAPER, R. M., 1988 Mechanisms of mutagenesis in the *Escherichia coli* mutator *mutD5:* role of mismatch repair. Proc. Natl. Acad. Sci. USA **85:** 8126–8130.

SELIGMAN, L.. and C. MANOIL, 1994 An amphipathic sequence determinant of membrane protein topology. J. Biol. Chem. **269:** 19888–19896.

STAHL, F., W. WENDE, A. JELTSCH and A. PINGOUD, 1996 Introduction of asymmetry in the naturally symmetric restriction endonuclease *Eco*RV to investigate intersubunit communication in the homodimeric protein. Proc. Natl. Acad. Sci. USA **93:** 6175–6180.

STEPHENS, K. M., R. J. MONNAT, JR., P. J. HEATH and B. L. STODDARD, 1997 Crystallization and preliminary X-ray studies of I-*Cre*I: a group I intron-encoded endonuclease from *C. reinhardtii.* Proteins: SFG **28:** 137–139.

THOMPSON, A. J., and D. L. HERRIN, 1991 *In vitro* self-splicing reactions of the chloroplast group I intron Cr.LSU from *Chlamydomonas reinhardtii* and *in vivo* manipulation via gene replacement. Nucleic Acids Res. **19:** 6611–6618.

THOMPSON, A. J., X. YUAN, W. KUDLICKI and D. L. HERRIN, 1992 Cleavage and recognition pattern of a double-strand-specific endonuclease (I-*Cre*I) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii.* Gene **119:** 247–251.

TURMEL, M., C. OTIS, V. CÔTÉ and C. LEMIEUX, 1997 Evolutionarily conserved and functionally important residues in the I-*Ceu*I homing endonuclease. Nucleic Acids Res. **25:** 2610–2619.

VILEJO, A. N., R. J. POGULIS and L.R. PEASE, 1995 Mutagenesis and synthesis of novel recombinant genes using PCR, pp. 603–612 in *PCR Primer: A Laboratory Manual,* Ed. 1, edited by C. W. DIEFFENBACH and G. S. DVEKSLER. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

WARING, R. B., R. W. DAVIES, C. SCAZZOCCHIO and T. A. BROWN, 1982 Internal structure of a mitochondrial intron of *Aspergillus nidulans.* Proc. Natl. Acad. Sci. USA **79:** 6332–6336.

Communicating editor: R. MAURER