Biochemical Studies of the Site-Specific I-*Ppo*I Endonuclease

by

Paula Kathleen Wittmayer

A dissertation submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

1996

BIOCHEMICAL STUDIES OF THE SITE-SPECIFIC I-*PPO*I ENDONUCLEASE

Paula Kathleen Wittmayer

Under the supervision of Associate Professor Ronald T. Raines

At the University of Wisconsin–Madison

I-*Ppo*I endonuclease, an intron-encoded endonuclease from the slime mold *Physarum polycephalum*, is a small enzyme (2 x 20 kDa) that catalyzes the cleavage of a large asymmetric DNA sequence (15 basepairs). This enzyme is a member of a relatively new class of endonucleases–homing endonucleases (Chapter 1). These enzymes "home" by initiating the propagation of their encoding sequence by creating a double-strand break within target DNA. In Chapter 2, the interactions of I-*Ppo*I with its substrate were examined during both binding (in the absence of Mg^{2+}) and catalysis (in the presence of Mg^{2+}). Using circular permutation assays, I-*Ppo*I was shown to induce a 38° ± 4° bend its substrate. Two independent methods, gel mobility shift assays and fluorescence polarization assays, revealed that I-*Ppo*I binds tightly to its substrate. Values of K_d range from 3.3 to 112 nM, increasing with increasing NaCl concentration. Similar salt effects on the values of K_m were observed during steady-state catalysis. At low salt concentrations, the value of k_{cat}/K_m for the cleavage of an oligonucleotide duplex approaches 10^8 $M^{-1}s^{-1}$. Although other divalent cations can replace Mg^{2+}, catalysis by I-*Ppo*I is most efficient in the presence of an oxophilic metal ion that can readily adopt an octahedral geometry: $Mg^{2+} > Mn^{2+} > Ca^{2+} = Co^{2+} > Ni^{2+} > Zn^{2+}$. In Chapter 3, the specificity of I-*Ppo*I is investigated by examining I-*Ppo*I catalysis with 92 mutant substrates that contain single or multiple basepair substitutions. I-*Ppo*I was found to tolerate some basepair degeneracy within its recognition sequence but not basepair

insertions or deletions. Multiple mutations within the recognition sequences were found to have a minimal, cumulative, or non-cumulative effect on catalysis by I-*Ppo*I. In addition, the sequences flanking the 15 basepair recognition site were shown to modulate recognition by I-*Ppo*I. In Chapter 4, site-directed mutagenesis was used to search for the active site residues of I-*Ppo*I. One residue, Lys142 was identified as an essential amino acid for catalysis. This study is the first detailed biochemical investigation of a homing endonuclease.

Acknowledgments

I am grateful for the countless people who have supported me over the duration of this work. Bob Lowery, Kim Knoche, and George Golumbeski supplied purified wild-type I-*Ppo*I. Rick Gourse and Wilma Ross provided the circular permutation vector, and Cathy Royer gave essential input on the fluorescence polarization studies (Chapter 2). Sung Hung Park helped overcome mishaps in creating mutant substrates, and Jennifer McKenzie contributed to the sequencing and assaying of these substrates (Chapter 3). Steve Mannino developed the expression vector and purification protocol for wild-type I-*Ppo*I which allowed for the study of variant proteins (Chapter 4). Jeung Hoi Ha guided me patiently through much of my first two years in a *bio*chemistry laboratory. Ruth Saecker aided me with many helpful discussions. Barbra Templer provided superior editing skills. All members of the Raines lab, both past and present, have been outstanding friends and co-workers. They have provided a pleasant environment in which to both work and play. Ron Raines provided good advice, patience, and support. I must also thank Lynn Zagzebski for her friendship and constant encouragement, my family for their patience, and Eric Eberhardt for easing all my frustrations and reminding me to have fun.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Considerable effort has been invested in detailed examinations of the molecular relationships between enzymes and their substrates. These studies have revealed that enzymes are exquisitely suited to catalyze specific reactions. Their catalytic proficiency can enhance rates of reactions up to 10^{17}-fold over the uncatalyzed reaction (Radzicka & Wolfenden, 1995). Their extraordinary specificity can be as precise as to distinguish atoms about a single chiral center (Walsh, 1979). Nonetheless, one area of enzymology where the enzyme–substrate relationship remains inadequately characterized involves the site-specific DNA nucleases. Much of the difficulty in delineating a chemical picture of the specificity and catalytic mechanisms of these proteins involves the polymorphic nature of their substrate, DNA; the incomplete understanding of the forces that modulate protein–DNA binding; and the absence of a model system that allows examination of both high substrate specificity and catalysis in a single protein. In an effort to develop such a unique model system, we have explored the physical properties of the interactions of the homing endonuclease I-*Ppo*I with its substrate (Chapter 2), the specificity of this enzyme (Chapter 3), and the amino acid residues critical for its catalytic efficiency (Chapter 4). These studies provide insight into the factors that govern site-specific DNA recognition and the mechanism of DNA cleavage by proteins.

1.2 Homing Endonucleases

A new class of enzymes has been discovered that provides a unique opportunity to examine substrate specificity and mechanisms of catalysis by enzymes. This unique class of enzymes consists of a group of endonucleases called homing endonucleases (Mueller *et al.*, 1993). These proteins catalyze the hydrolysis of one or two phosphodiester bonds within a specific sequence of double-stranded DNA. The activity of homing endonucleases results in the generation of 2 to 4 base overhangs, depending on the

individual enzyme (Dujon, 1989; Perlman & Butow, 1989; Mueller *et al.*, 1993). Therefore, *in vitro,* the activity of homing endonucleases is similar to that of the familiar restriction enzymes: site-specific cleavage of DNA. A significant difference, however, is the size of their recognition sites. Homing endonucleases have recognition sequences ranging in size from 15 to 40 basepairs as compared to the 4 to 8 basepairs recognized by typical restriction enzymes (Mueller *et al.*, 1993; Roberts & Halford, 1993). Unlike restriction sites, the large recognition sequences of homing endonucleases are usually asymmetric. Homing endonucleases, therefore, appear to have achieved a more intricate and extensive specificity while still maintaining catalytic efficiency. The extraordinary specificity of homing endonucleases makes them attractive systems for examining both complex recognition and cleavage of DNA.

Homing endonucleases exist as open reading frames (ORFs) located within either group I introns or archaeal introns, or as inframe spacers [inteins, (Perler *et al.*, 1994)] within protein-coding sequences (Lambowitz & Belfort, 1993). They function to promote the self-propagation of their encoding region, a process called intron or intein homing. The endonuclease targets an intronless or spacerless allele for conversion into an intron-containing or spacer-containing allele (Dujon, 1989; Lambowitz & Belfort, 1993). When an intron- or intein-plus strain of an organism is crossed with an intron- or intein-minus strain, the resulting progeny contain the intron or intein within each allele. A general schematic of homing is shown in Figure 1.1. The encoded homing endonuclease initiates the homing process by site-specific cleavage of the recipient, intronless or inteinless allele (Dujon, 1989). The creation of the double-stranded break allows for duplication of the intron or intein into the recipient allele. This recombination event results in two copies of the intron or intein, with one copy present in each allele. The great specificity and efficiency of homing (nearly 100% conversion of minus alleles to plus alleles) make it an extraordinary recombinatory event (Dujon, 1981).

Intron homing was first described genetically over twenty-five years ago, ironically even before the discovery of introns (Coen *et al.*, 1970). Crosses of yeast strains showed that certain mutations displayed non-reciprocal recombination and a highly biased transmission of alleles. The alleles of one parent were over-represented in the progeny at the expense of the other parent's alleles. The unique genetic element dictating this biased recombination was later identified as the ω intron, a 1.1 kb intron found in the large (21S) rRNA gene of the mitochondrial DNA of some, but not all, strains of *Saccharomyces cerevisiae*. (Bos *et al.*, 1978; Faye *et al.*, 1979; Dujon, 1980). The ω intron contains an open reading frame for a 30 kilodalton (kDa) polypeptide which was identified to be a site-specific endonuclease (Colleaux *et al.*, 1986). Site-specific DNA cleavage activity of the intron-encoded protein was demonstrated and provides a basis for the site-specific and efficient nature of the ω intron's mobility (Colleaux *et al.*, 1986).

Intron homing is characteristically unidirectional and efficient. Crosses between ω^+ and ω^- strains of *Saccharomyces cerevisiae* revealed that the ω^+ allele was propagated nearly 100% whereas the ω^- allele was nearly depleted from the population (Dujon, 1981). Included with the propagation of the intron was the co-conversion of flanking regions; the frequency of co-conversion varied inversely with the distance of the nucleotides from the intron borders (Jaquier & Dujon, 1985). Further investigation of young zygotes from the crosses of ω^+ and ω^- strains of yeast revealed the formation of a double-stranded break in the intronless gene (Dujon *et al.*, 1985). This break disappeared within hours of formation. These observations lead to the proposal that intron homing was mediated by a double-stranded-break repair mechanism (Szostak *et al.*, 1983).

1.3 Double-Strand-Break Repair Model

The double-strand break repair model of recombination was first proposed by Rothstein and coworkers (Szostak *et al.*, 1983). This model is used to explain the genetic

switching that occurs with yeast mating types. An important feature of this model is the creation of a double-strand break or gap. The double-strand gap is repaired by two rounds of single-strand synthesis followed by the resolution of the two generated Holiday junctions. A scheme of their model is shown in Figure 1.2. Recombination is initiated by the creation of a double-strand break in the recipient DNA duplex. The cut is enlarged to create a gap by exonucleases followed by invasion of the homologous region of the donor duplex by the newly freed 3′ strand. The invading strand acts as a primer to initiate single-strand synthesis with the donor strand acting as a template. Repair synthesis continues, displacing a larger portion of the complementary strand of the donor duplex until the displaced region contains complementary sequence to the 3′ end of the other side of the gapped region. This displaced single strand can then anneal and act as a primer for a second round of repair in the recipient duplex. Resolution of the recombination junctions creates two DNA duplexes, both containing the homologous region (intron or intein).

Recombination experiments performed in yeast with hybrid plasmids (*Escherichia coli* plasmids containing fragments of yeast DNA) support this model (Orr-Weaver *et al.*, 1983). Rothstein and coworkers showed that recombination increased up to 3000-fold when a double-stranded break was introduced into the yeast region of the transforming plasmid by previous treatment of the plasmid with a restriction enzyme (Orr-Weaver *et al.*, 1981). They also showed that the region of homologous recombination could be "targeted" to recombine at only one of the multiple regions of homogeneity by creating a double-strand break within the desired region of homology (Orr-Weaver *et al.*, 1981). In addition, if the hybrid plasmid was treated with two restriction enzymes, thereby containing a double-stranded gap within the yeast region, the gapped region was always repaired with genomic DNA sequence and never resulted in deletions. The repair of the gapped region with the genomic DNA sequence occurs with high fidelity as demonstrated

by the efficient regeneration of a functional gene from a gapped-gene plasmid (Orr-Weaver *et al.*, 1981). Finally, this process of recombination requires the product of the RAD52 gene. This gene is known to be needed for the repair of double-stranded breaks (Resnick & Martin, 1976).

Collectively, the evidence points towards the necessity of a double-strand break for the recombination event to occur. The double-stranded break model is consistent with observations of intron homing: efficient and unidirectional homing of the intron, observable creation of a double-stranded break that disappears with time, and co-conversion of sequences flanking the intron (Dujon, 1989). For the process of intron or intein homing, the homing endonucleases create this double-strand break within the recipient DNA duplex. All homing endonucleases investigated, except for two, cleave DNA at a specific recognition sequence creating double-strand breaks (Mueller *et al.*, 1993). The two exceptions, I-*Hmu*I and I-*Hmu*II (for nomenclature see Figure 1.3), appear to promote exclusion of the *Bacillus subtilis* phages SPO1 and SP82 (Goodrich-Blair & Shub, 1996). Both enzymes, I-*Hmu*I and I-*Hmu*II specifically nick double-stranded DNA *in vitro*, and the fate of the nick *in vivo* is under investigation (Goodrich-Blair & Shub, 1996). The recognition sites and cleavage patterns of homing endonucleases are illustrated in Figure 1.3. A recognition site is comprised of sequences from both the upstream and downstream exons that surround the intron insertion site (Dujon *et al.*, 1989). Successful intron or intein homing disrupts the recognition sites of the homing endonuclease thereby preventing further cleavage of the organism genome.

1.4 Location of Homing Endonucleases

The sources of homing introns and their endonucleases are diverse as seen in Figure 1.3. Examples have been found in each of the three kingdoms: archaea, eubacteria, and eukarya (Mueller *et al.*, 1993). The genetic environments in which the homing

endonucleases are located range from mitochondrial, chloroplast, nuclear, to bacteriophage DNA. The original discovery of the ω intron-encoded endonuclease, now termed I-*Sce*I, was thought to be an anomaly of the yeast ω system (Jaquier & Dujon, 1985). But soon there after, the discovery of I-*Sce*II, I-*Ppo*I, I-*Tev*I, and I-*Tev*II from *Saccharomyces cerevisiae*, *Physarum polycephalum*, and the T-even phages, respectively, revealed the more general nature of these homing introns and endonucleases (Bell-Pedersen *et al.*, 1989; Delahodde *et al.*, 1989; Quirk *et al.*, 1989; Bell-Pedersen *et al.*, 1990; Muscarella *et al.*, 1990). Reports of other intron-encoded endonucleases rapidly followed with several of the *Chlamydomonas* organelles containing homing endonucleases: I-*Csm*I, I-*Cre*I, I-*Ceu*I, I-*Chu*I (Colleaux *et al.*, 1990; Dürrenberger & Rochaix, 1991; Gauthier *et al.*, 1991; Cote *et al.*, 1993). When the homing endonuclease I-*Dmo*I from an archaeal intron of *Desulfurococcus mobilis* was announced, the taxonomic range of the intron-ecoded endonucleases expanded (Dalgaard *et al.*, 1993). The existence of intron-encoded endonucleases in both group I and archaeal introns ignited the debate on the ancestry of homing introns.

Further complexities in the classification of homing endonucleases developed when an even more remarkable source of homing endonucleases was identified: protein splicing (Kane *et al.*, 1990). Protein splicing was first identified with the gene product of the VMA1/TFP1 gene in *Saccharomyces cerevisiae*. Here, it was revealed that the entire 119 kDa polypeptide is translated as a precursor protein containing the amino acid sequence of two full proteins. One protein, the "intein," separates the two "exteins" of the second protein. After translation, the intein is excised subsequently via protein splicing, and the two exteins are fused (Kane *et al.*, 1990). The splicing results in the formation of the functional 69-kDa ATPase and the 50-kDa site-specific endonuclease PI-*Sce*I (for nomenclature see Figure 1.3 (Perler *et al.*, 1994)) (Kane *et al.*, 1990; Gimble & Thorner, 1992). As with intron homing, PI-*Sce*I recognizes and cleaves an inteinless VMA1 gene,

resulting in the homing of the endonuclease coding sequence (Gimble & Thorner, 1992). Similar systems were uncovered in DNA polymerases of *Thermococcus litoralis* (Hodges *et al.*, 1992) and *Pyrococus sp.* strain GB-D (Perler *et al.*, 1992), and the RecA protein of *Mycobacterium tuberculosis* (Dujon *et al.*, 1989; Davis *et al.*, 1992), yielding PI-*Tli*I, PI-*Psp*I, and PI-*Mtu*I endonucleases respectively (Perler *et al.*, 1994).

The diverse locations and genomic surroundings of the various homing endonucleases lead to a variety of expression methods. Homing endonucleases can be translated from a precursor RNA transcript (such as I-*Sce*I (Zhu *et al.*, 1987)), from an independent RNA transcript (such as I-*Tev*I and I-*Tev*II (Gott *et al.*, 1988)), or from a processed RNA transcript (such as I-*Dmo*I (Dalgaard *et al.*, 1993)). The gene can be free-standing or inframe with the upstream exon. The free-standing genes are translated independently of the exons such as for the endonucleases I-*Tev*I and I-*Tev*II (Gott *et al.*, 1988). The majority of the known homing endonucleases are inframe with the upstream exon and are translated as a fused polypeptide such as for the endonuclease I-*Sce*II (Wernette *et al.*, 1990). The fused polypeptide is then subsequently processed to produce the mature endonuclease. Integral to the method of homing endonuclease expression is the level at which the encoding intron or intein splices.

1.5 Intron and Intein Splicing

Group I introns are a class of mobile genetic elements capable of self-splicing at the RNA level (Cech, 1990). They have characteristic splicing mechanisms and structural features. Group I introns splice via two transesterification steps: the first involving the addition of a guanosine or guanosine nucleotide to the 5′ end of the intron where the nucleophile is the 3′ hydroxyl of the exogenous guanosine, the second involving attack on the phosphorous at the 3′ splice site by the newly freed 3′ hydroxyl of the 5′ exon (Zaug *et al.*, 1983; Cech, 1987). These reactions result in the ligation of the exons and excision

of the intron. Both steps are catalyzed by RNA *in vitro*, yet the splicing of some introns may be modulated *in vivo* by accessory factors, proteins independent from the intron (Burke, 1988; Lambowitz & Perlman, 1990). Group I introns vary in length from 200-3000 nucleotides. Their consensus structure has been elucidated through structure prediction methods and comparative sequence analysis in combination with structural, enzymatic, and chemical probes (Davies *et al.*, 1982; Michel *et al.*, 1982; Cech *et al.*, 1983; Inoue & Cech, 1985; Tanner & Cech, 1985). Basic features of group I introns are shown in Figure 1.4. The core region is composed of a common set of paired regions (P1-P9). Extra sequences form stem loops that are peripheral to the core. These regions are more variable between different introns and are the locations of the optional homing endonuclease open reading frames (ORFs) (Burke, 1988; Cech, 1988). Association of a homing endonuclease ORF with the peripheral regions prevents interference with catalysis by the ribozyme. Some ORFs do, however extend beyond the loop regions and into the intron core (Gott *et al.*, 1988; Delahodde *et al.*, 1989; Muscarella *et al.*, 1990; Turmel *et al.*, 1991). These sequences have adapted to allow both intron splicing and translation of the functional endonuclease.

Archaeal introns from halophiles and thermophiles splice to form mature gene products of tRNA and rRNA genes (Thompson & Daniels, 1988; Thompson & Daniels, 1990). Their sizes vary from 15 nucleotides up to approximately 600 nucleotides. These introns are not autocatalytic; rather, they rely on an endoribonuclease for excision (Kjems & Garrett, 1991). The endoribonuclease recognizes a bulge-helix-bulge motif located at the borders of the archaeal introns and cleaves the polynucleotide producing 5′-OH and 2′,3′-cyclic phosphodiester ends (Kjems & Garrett, 1991). The exons are subsequently ligated to form the mature RNA gene product. Because there is no ribozyme catalytic core with archaeal introns, the intron sequence is quite variable aside from the bulge-

helix-bulge motif. Likewise, the positioning of homing endonuclease ORFs within the intron is presumably less critical.

Protein or intein splicing shares similarities with group I intron splicing. The autocatalytic process excises the intervening protein sequence (intein) and ligates the external sequences (exteins) to form two functional protein products (Cooper *et al.*, 1993). Protein splicing breaks and reforms a peptide bond, whereas group I intron splicing breaks and reforms a phosphodiester bond (Cech, 1987; Cooper *et al.*, 1993). Unlike group I intron splicing, the precise mechanism of protein splicing is not yet known; however, the pool of information gathered on this process is growing. Structurally, all the information needed for protein splicing is contained within the intein. For example, when the spacer region encoding PI-*Sce*I is fused into a foreign context, the VAT2 gene, the intein is still able to splice and produce two functional proteins (Cooper *et al.*, 1993). Similarly, the PI-*Mtu*I intein excises successfully from β-galactosidase when the endonuclease coding sequence, together with 8 N-terminal and 32 C-terminal flanking codons, is inserted into the *LacZ* gene (Davis *et al.*, 1992).

The amino acid sequences of the inteins are relatively long, encoding polypeptides of 360 to 537 residues with little sequence identity except at the intein-extein boundaries (Cooper & Stevens, 1995). The first residue of the intein sequence (the N-extein-intein boundary) is a thiol- or hydroxyl-containing residue (Cys, Thr, or Ser). The tripeptide sequence His-Asn-(Cys/Ser/Thr) comprises the C-extein-intein border (see Figure 1.5). Mutation of these residues is deleterious to protein splicing (Hodges *et al.*, 1992; Cooper *et al.*, 1993). For example, the invariant asparagine residue is absolutely required as substitution by the chemically similar residue glutamine prevents splicing (Cooper *et al.*, 1993).

Two additional discoveries have lead to insights for the chemical mechanism of protein splicing: first, the identification of an intermediate along the splicing pathway;

and second, the identification of the covalent structure of the carboxyl terminus of the excised intein (Xu *et al.*, 1993; Xu *et al.*, 1994). The first discovery showed that a branched polypeptide intermediate forms during splicing (Xu *et al.*, 1993). Using a hybrid precursor protein expressed in *Escherichia coli*, MIP (a fusion protein consisting of the *M*altose-binding protein as the N-extein, the *I*ntein from *Pyrococcus sp* pol gene, and *P*aramyosin as the C-extein), an intermediate species along the splicing pathway was observed with a slower migration on sodium dodecyl sulfate-polyacrylamide gels than that of the precursor protein (Xu *et al.*, 1993). Amino-terminal sequencing of the intermediate showed two amino termini, one corresponding to the N-extein and the other to the intein. These results demonstrated that the intermediate was linked covalently to form a branched polypeptide. The second discovery demonstrated that the C-terminus of the intein sequence is composed of an aminosuccinimide residue (Xu *et al.*, 1994; Shao *et al.*, 1995). Analysis of the C-terminal cyanogen bromide peptide of PI-*Psp*I suggests that the terminal asparagine residue cyclizes to form a succinimide ring product. Asparagine residues have a propensity to undergo rearrangements resulting either in deamidation of the amino acid side chain or cleavage of the peptide bond via formation of aminosuccinimide (see Figure 1.6) (Voorter *et al.*, 1988). This information, in conjunction with the mutational analysis of the conserved asparagine residue, suggests that the branched intermediate (above) is resolved by the cyclization of the conserved asparagine residue into the two protein products–the functional extein and the intein homing endonuclease.

1.6 Specificity of Homing Endonucleases

A distinctive feature of homing endonucleases is their specificity. Their recognition sites are lengthy, approximately 15 to 40 basepairs, and typically asymmetrical (Perlman & Butow, 1989). The stringency of recognition varies. Some homing endonucleases

require near complete compliance with the recognition sequence while others allow for levels of degeneracy. For example, the enzyme I-*Sce*I is highly specific; this endonuclease cleaves less than once in 10,000 kilobases (Thierry *et al.*, 1991). Fourteen single basepair substitutions within its 18 basepair recognition sequence severely affect activity by the enzyme and an additional eleven single basepair substitutions affect the activity moderately (Colleaux *et al.*, 1988). The endonuclease I-*Ceu*I also maintains a high level of specificity. Eleven single mutations within its 19 basepair recognition sequence completely eliminate cleavage by I-*Ceu*I (Marshall & Lemieux, 1992).

The enzyme I-*Sce*II is less stringent; this enzyme cleaves DNA approximately once in every 100 kilobases (Wernette *et al.*, 1992). Of the 36 single basepair substitutions examined for this homing endonuclease, most were tolerated by the enzyme. For example, only nine mutations blocked cleavage moderately and only three blocked cleavage completely (Wernette *et al.*, 1992). Limited analysis of the I-*Cre*I endonuclease reveals that this enzyme cleaves its recognition site with a level of specificity similar to I-*Sce*II (Dürrenberger & Rochaix, 1993). Single and multiple mutations are tolerated by the enzyme. Moreover, DNA with a single A to G mutation at the border of the recognition site appears to be a better substrate for I-*Cre*I than is its natural substrate (Dürrenberger & Rochaix, 1993).

The more permissive endonucleases are the T4 phage endonucleases, I-*Tev*I and I-*Tev*II. These enzymes cleave once every 10 kilobases (Quirk *et al.*, 1989; Chu *et al.*, 1991). No single basepair within the 48 basepairs spanning the homing site of I-*Tev*I is essential for cleavage (Byrk *et al.*, 1993). Interestingly, both phage enzymes, I-*Tev*I and I-*Tev*II, make a majority of their contacts with the minor groove, consistent with the minor groove providing fewer complementary basepair functional groups for sequence determination (see Chapter 3) (Byrk *et al.*, 1993; Loizos *et al.*, 1996).

1.7 Structural Motifs of Homing Endonucleases

The homing endonucleases contain three major structural families based on amino acid sequence homology: the LAGLI-DADG motif, the GIY-YIG motif, and the H-N-H motif (Figure 1.7) (Michel *et al.*, 1982; Michel & Dujon, 1986; Gorbalenya, 1994; Shub *et al.*, 1994). The majority of the homing endonucleases belong to the LAGLI-DADG family (Mueller *et al.*, 1993). Representatives of this family of endonucleases are present in all three kingdoms, from both introns and inteins. This motif contains two amino acid repeats with the sequence LAGLI-DADG encompassed within 12 amino acids. These repeats are referred to as dodecapeptides, or P1 and P2. The two repeats are separated by approximately 100 amino acid residues.

The LAGLI-DADG motif is also found in proteins that are not homing endonucleases. Initial identification of the dodecapeptide motif occurred in yeast mitochondrial maturases, enzymes which are involved in RNA processing (Michel *et al.*, 1982; Waring *et al.*, 1982; Hensgens *et al.*, 1983). In addition, the motif appears in two non-intron-encoded endonucleases: the yeast mitochondrial nuclease, Endo-*Sce*I (Nakagawa *et al.*, 1991), and the HO endonuclease involved in yeast mating-type switching (Russell *et al.*, 1986). DNA cleavage activity has been linked to amino acids within the motif via mutational analysis of the conserved repeats. Mutation of the conserved acidic residue (LAGLI-DA<u>D</u>G) in the first repeat (or P1) of the endonuclease PI-*Tli*I results in the loss of DNA cleavage activity (Hodges *et al.*, 1992). Mutations of the conserved acidic residues of both repeats (P1 and P2) for the endonuclease PI-*Sce*I also prevent DNA cleavage activity but leave DNA binding activity intact (Gimble & Stephens, 1995). Alteration of both glycine residues of the P1 repeat (but not the P2 repeat) of the endonuclease I-*Sce*II prevented intron homing and endonucleolytic activity (Henke *et al.*, 1995). The dodecapeptide motif is especially prevalent in the homing endonucleases expressed as either inteins or as proteins fused inframe with upstream

exons (Section 1.4). This observation has lead to speculation that the dodecapeptide may have a role in proteolytic as well as nucleolytic activity; however, proteolytic activity has not been demonstrated (Shub & Goodrich-Blair, 1992).

The second motif GIY-YIG is another consensus sequence found in homing endonucleases (Michel & Dujon, 1986). The two components of the amino acid motif are separated by 10 or 11 residues. The homing endonuclease I-*Tev*I contains this motif as do the open reading frames of a number of group I introns found in fungal mitochondria (Burger & Werner, 1985). In addition, this motif is found in some non-intron-encoded proteins of bacteriophage T4 such as the site-specific endonuclease SegA (Sharma *et al.*, 1992).

The final motif, H-N-H, contains a consensus sequence spanning 30-33 amino acids with two highly conserved pairs of histidine residues (EX_1HH-HX_3H) (Gorbalenya, 1994; Shub *et al.*, 1994). The H-N-H motif is found in the homing endonucleases from bacteriophages of *Bacilllus subtilis* (I-*Hmu*I) and *Escherichia coli* (I-*Tev*III) (Goodrich-Blair *et al.*, 1990; Eddy & Gold, 1991). This motif is also seen in a number of other bacterial endonucleases (McrA and colicins) and the zinc finger-like domain of group II open reading frames (Gorbalenya, 1994; Shub *et al.*, 1994). The association of this motif with zinc finger-like domains (CXXC) has lead to the proposal that the C-terminal HXXXH component may function as an additional "zinc finger" to coordinate metal ions, with the remainder of the H-N-H motif forming the active site (Shub *et al.*, 1994). However, no experimental evidence has been presented to support this idea.

1.8 Origins of Homing Endonucleases

Homing endonucleases are distributed sporadically throughout the three kingdoms with no apparent evolutionary lineage (Dujon, 1989). For example, various homing endonucleases contain similar peptide motifs such as LAGLI-DADG (Figure 1.7), yet

they are found in phylogenically distant organisms. The homing endonucleases I-*Tev*I and I-*Tev*II are not homologous and have different insertion sites despite being located in two related T4 bacteriophage introns that share 60% sequence identity (Shub *et al.*, 1988). The introns found in *Physarum polycephalum* and *Tetrahymena* share 70% sequence identity and identical positioning within the rRNA gene; however, only the *Physarum polycephalum* intron contains a homing endonuclease (Muscarella & Vogt, 1989). The occurrence of these homing enzymes appears to be haphazard.

The molecular evolution of homing endonucleases is intriguing. In addition to their widespread nature, the endonucleases have different modes of expression, are found in variable genetic environments, and have irregular positioning of their respective insertion sites (Dujon, 1989; Mueller *et al.*, 1993). Attempts to draw lines of descent for homing endonucleases have been challenging (Doolittle, 1993). Two theories have emerged that represent evolutionary extremes. The first suggests that there existed an early ancestral homing endonuclease that predates the division of the three kingdoms. The second proposes that homing endonucleases evolved later, and their mobility allowed horizontal transfer across kingdom lines (Dujon, 1989; Dalgaard *et al.*, 1993).

A common theme that unites many of the evolution hypotheses is that homing endonuclease ORFs have evolved independently from introns (Lambowitz & Belfort, 1993). The endonuclease coding sequence itself is thought to be the invasive and selfish DNA unit, its propagation dependent on the successful colonization of genetically neutral sites (Belfort, 1989; Lambowitz, 1989; Shub & Goodrich-Blair, 1992). Introns provide such a hospitable, neutral landing site for an ORF. In return, the intron becomes mobile itself after invasion by a homing endonuclease. This synergistic relationship between the intron and ORF results in intron homing. The fortuitous landing of an ORF within a region capable of self-splicing at the protein or RNA level would also allow incorporation of the endonuclease coding sequence into the genome. Splicing would maintain host

viability by preventing disruption of essential genes, yet still allow for ORF propagation. As a result, the ORF becomes a successful parasite. Such mobility events could account for the presence of extremely similar homing endonucleases in widely diverse biological niches (Dujon, 1989; Mueller *et al.*, 1993). While the debates continues on the ancestry of these enzymes, it is provocative to envision a scenario of mobile endonucleases distributing themselves throughout biological genomes with differing evolutionary fates.

1.9 I-*Ppo*I and the Pp LSU 3 Intron

The homing endonuclease at the focus of this study, I-*Ppo*I, is derived from the ORF found within intron 3 of the gene encoding the large ribosomal subunit of the slime mold, *Physarum polycephalum.* Intron 3 (Pp LSU 3) is an optional intron found only in the Carolina strain of *Physarum polycephalum* (Muscarella & Vogt, 1989; Muscarella *et al.*, 1990). Crosses of the Carolina strain (intron-plus) with other *Physarum* strains (intron-minus) results in the conversion of the rDNA lacking the intron to intron-containing alleles. I-*Ppo*I mediates this homing event by recognizing and cleaving the DNA duplex lacking the Pp LSU 3 intron.

Intron 3 is one of three group I introns found within the nuclear 26S ribosomal gene. This intron is positioned within the rDNA gene in the exact position as the well-studied *Tetrahymena* intron from *Tetrahymena thermophila.* The two introns, Pp LSU 3 and Tt LSU 1, are closely related sharing 70% sequence identity excluding the extra open reading frame found within the slime-mold intron (Muscarella & Vogt, 1989). The splicing reaction of the solitary *Physarum* intron, however, is 1500-fold slower than that of the *Tetrahymena* intron (Rocheleau & Woodson, 1994). The presence of the second intron (Pp LSU 2) 24 nucleotides downstream of intron 3 enhances splicing of intron 3 10-25 fold (Rocheleau & Woodson, 1995). In addition to the typical self-spicing mechanism of group I introns (Section 1.5), the *Physarum* intron undergoes a secondary

reaction (Ruoff *et al.*, 1992). In this reaction, another guanosine adds to an internal processing site located after the ORF, cleaving the intron into two RNA fragments. The 5′ product contains the ORF of I-*Ppo*I; the 3′ product, the entire ribozyme catalytic core.

Expression of I-*Ppo*I from the ORF within the Pp LSU 3 intron can be initiated at either of two start codons yielding polypeptides of 160 and 183 amino acids (Muscarella *et al.*, 1990). The AUG start codon of the longer species is positioned in the upstream exon while the coding sequence for the shorter species remains entirely within the intron. Both proteins are active *in vitro* and contain no apparent DNA-binding or homing-endonuclease amino-acid motifs (Figure 1.7). In the presence of Mg^{2+}, the activity of I-*Ppo*I creates a double-strand break within a 15 basepair recognition sequence as determined by deletion analysis (Figure 1.8) (Ellison & Vogt, 1993). The cleavage event generates four-base overhangs leaving 3′-hydroxyl groups. The recognition sequence encompasses both the four-base cleavage site and the intron insertion site. Ten of the fifteen basepairs within this recognition site contribute to an imperfect palindrome. This pseudosymmetry contrasts I-*Ppo*I with other homing endonucleases; most recognition sites of homing endonuclease contain no elements of symmetry (Figure 1.3) (Perlman & Butow, 1989). The I-*Ppo*I endonuclease functions as a dimer consistent with the partial symmetry of the recognition site (Ellison & Vogt, 1993). The I-*Ppo*I footprint with the small chemical probe dimethyl sulfate (DMS) is, however, strikingly asymmetric (Ellison & Vogt, 1993). The protein contacts its DNA substrate over a 22-24 basepair stretch with differing patterns of protection about the pseudo symmetric halves of the recognition site (Ellison & Vogt, 1993).

The homing endonuclease I-*Ppo*I has several characteristics that make the enzyme useful as a model system. First, I-*Ppo*I has a fairly small size (2 x 20 kDa) with a large recognition sequence (15 basepairs). I-*Ppo*I has a high catalytic efficiency and relative stability compared to other homing endonucleases (Lowery *et al.*, 1992). Finally, the

existence of a bacterial expression system makes I-*Ppo*I an ideal system to investigate the enzymology and specificity of homing endonucleases (S. J. Mannino and R. T. Raines, unpublished results). Investigations of I-*Ppo*I will lead to new insights on the protein–DNA interactions and site-specific catalysis of homing endonucleases. Compared to other endonucleases such as restriction endonucleases, or other DNA–binding proteins such as bacterial transcription factors, little is known about this dynamic family of proteins. In this study, we look at the physical properties of the homing endonuclease I-*Ppo*I and its interactions with its substrate (Chapter 2), the specificity of the enzyme (Chapter 3), and the amino acid residues critical for its catalytic activity (Chapter 4).

Figure 1.1 Intron homing. The open box and solid black lines represent the intron and exons, respectively. N indicates the basepairs of the endonuclease recognition sequence. The homing endonuclease recognizes its homing site (recognition site) in the intron-minus allele and creates a double-strand break. The generation of this break initiates the transfer of a copy of the intron into the recipient allele resulting with both alleles containing the intron.

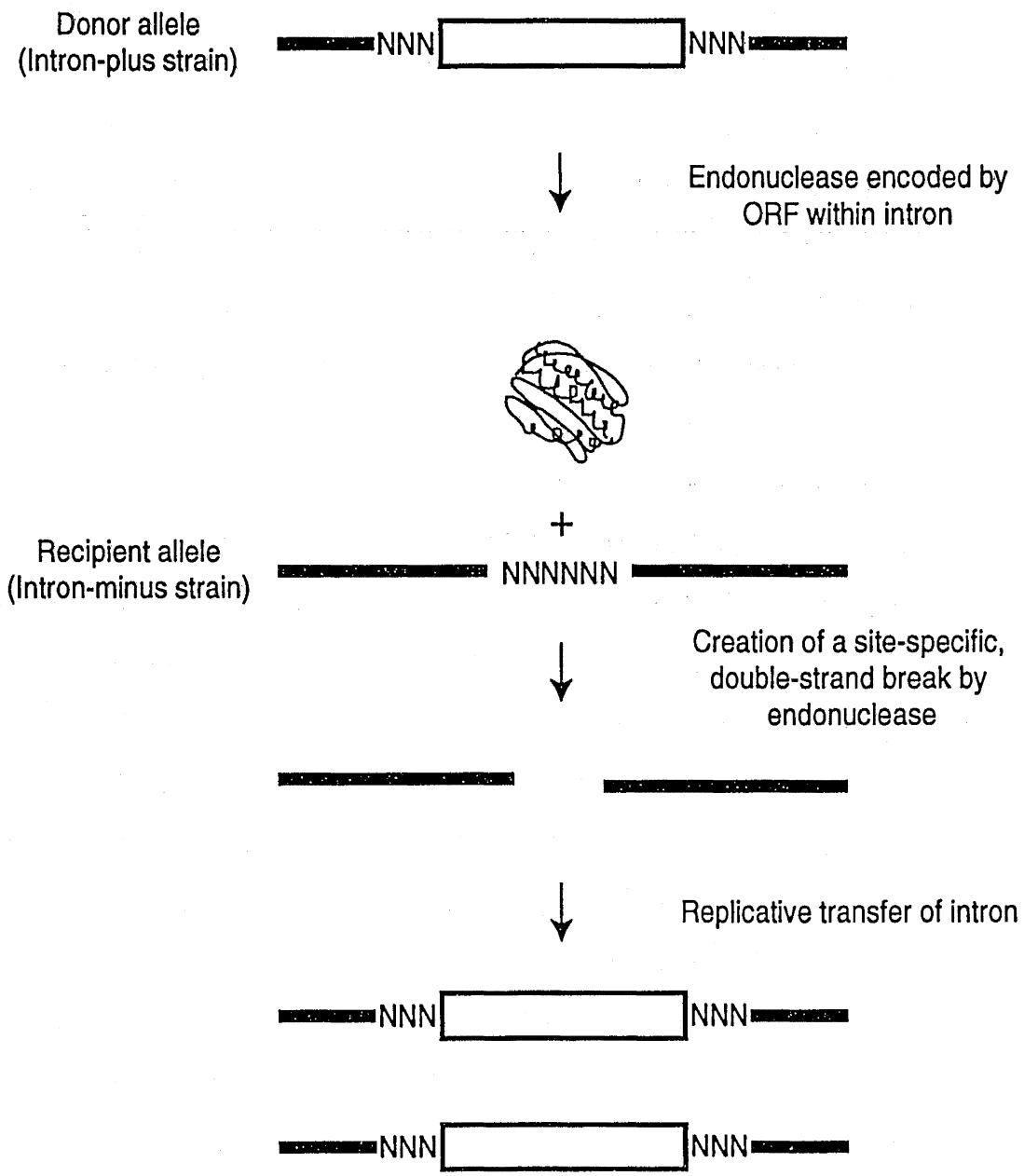

Donor allele
(Intron-plus strain)
NNN
NNN
Endonuclease encoded by
ORF within intron
+
Recipient allele
(Intron-minus strain)
NNNNNN
Creation of a site-specific,
double-strand break by
endonuclease
Replicative transfer of intron
NNN
NNN
NNN
NNN

Figure 1.2 Double-strand break model (DSBM) of Szostak *et. al.* (1983). The thin and thick lines represent the intron donor and intron acceptor, respectively. The dashed lines indicate newly synthesized DNA. The homing endonuclease creates the initial break in the recipient DNA followed by exonucleolytic degradation. The freed 3′-OH strand invades the donor duplex and initiates a first round of single-strand synthesis followed by a second round of single-strand synthesis. Resolution of the recombination junctions results in the duplication of the intron and flanking sequences into the recipient DNA duplex.

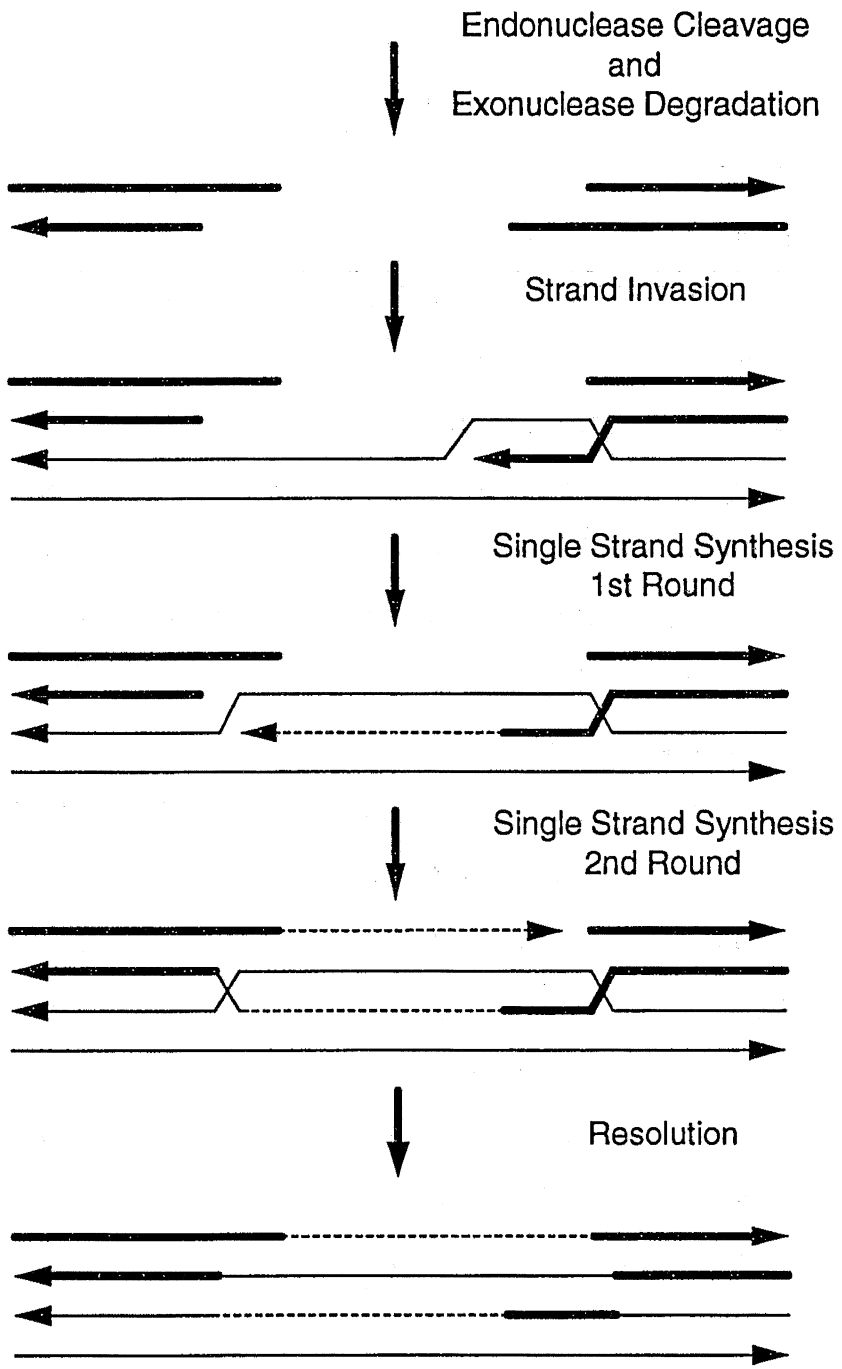

Endonuclease Cleavage
and
Exonuclease Degradation
Strand Invasion
Single Strand Synthesis
1st Round
Single Strand Synthesis
2nd Round
Resolution

Figure 1.3 Homing site sequences of homing endonucleases. The intron or intein insertion site is indicated by the space in the sequence denoted with an arrow. The cleavage sites, if known, are indicated by the staggered line. The endonucleases named with the prefix I are Intron-encoded, and the endonucleases with the prefix PI are generated by protein splicing or Protein Inserts. The following three letter abbreviation represents the *Genus-species* of the organism from which the endonuclease is found. The roman numeral suffix designates the chronological order of the enzyme's discovery (Dujon *et al.*, 1989; Perler *et al.*, 1994). This compilation has been updated from Lambowitz & Belfort (1993).

Homing Endonuclease	Source	Homing Site
Eukarya		
I-*Sce*I	*Saccharomyces cerevisiae*	TAGGGAT AACAGGGTAAT ATCCCTA TTGTCCCATTA
I-*Sce*II	*Saccharomyces cerevisiae*	TTGGT CACCCTGAAGTAT AACCA GTGGGACTTCATA
I-*Sce*III	*Saccharomyces cerevisiae*	GGTTTTGGT AACTATTTATTAC CCAAAACCA TTGATAAATAATG
PI-*Sce*I	*Saccharomyces cerevisiae*	ATCTATGTCGGGTGC GGAGAAAGAGGTAAT TAGATACAGCCCACG CCTCTTTCTCCATTA
I-*Ppo*I	*Physarum polycephalum*	CTCTCT TAAGGTAGC GAGAGA ATTCCATCG
I-*Csm*I	*Chlamydomonas smithii*	GTACTAGCATGGGGT CAAATGTCTTTCTGG CATGATGGTACCCCA GTTTACAGAAAGACC
I-*Cre*I	*Chlamydomonas reinhardtii*	CAAAACGT CGTGAGACAGTTTGGT GTTTTGCA GCACTCTGTCAAACCA
I-*Ceu*I	*Chlamydomonas eugametos*	TCTAACTATAACGGT CCTAAGGTAGCGAGG AGATTGATATTGCCA GGATTCCATCGCTCC
I-*Chu*I	*Chlamydomonas humicola*	GAAGGTTTGGCACCT CGATGTCGGCTCATC CTTCCAAACCGTGGA GCTACAGCCGAGTAG
I-*Cpa*II	*Chlamydomonas palldostigmatica*	TCCGGCT AACTCTGTGCCAG AGGCCGA TTGAGACACGGTC
Archaea		
I-*Dmo*I	*Desulfurococcus mobilis*	GCCTTGCCGGGTA AGTTCCGGC CGGAACGGCCCAT TCAAGGCCG
PI-*Tli*I	*Thermococcus litoralis*	GTTCTTTATGCGGAC ACTGACGGCTTTTAT CAAGAAATACGCCTG TGACTGCCGAAAATA
PI-*Tli*II	*Thermococcus litoralis*	AAATTGCTTGCAAAC AGCTATTACGGCTAT TTTAACGGACGTTTG TCGATAATGCCGATA
PI-*Psp*I	*Pyrococcus species GB-D*	AAAATCCTGGCAAAC AGCTATTATGGGTAT TTTTAGGACCGTTTG TCGATAATACCCATA
I-*Por*I	*Pytobaculum organotrophum*	GAGCCCGTA AGGGTGTGTAC CTCGGGCAT TCCCACACATG
Eubacteria		
I-*Tev*I	*Bacteriophage T4 (T-even phage)*	AACGCTCAGTAGATGTTTTCTTGGGT CTACCGTTTAA TTGCGAGTCATCTACAAAAGAACCCA GATGGCAAATT
I-*Tev*II	*Bacteriophage T4 (T-even phage)*	GCTTATGAGT ATGAAGTGAACACGTTATTC CGAATACTCA TACTTCACTTGTGCAATAAG
I-*Tev*III	*Bacteriophage T4 (T-even phage)*	GTTTTTATGTATCTTTTGCGT GTACCTTTAACTTC CAAAAATACATAGAAAACGCA CATGGAAATTGAAG

Figure 1.4 Structure of the intron ribozyme (Mueller *et al.*, 1993). The location of endonuclease ORFs are shown within the model of group I introns. The filled boxes represent the exons. The dashed lines indicates long-range tertiary interactions. Asterisks indicate ORFs which extend into the intron stems.

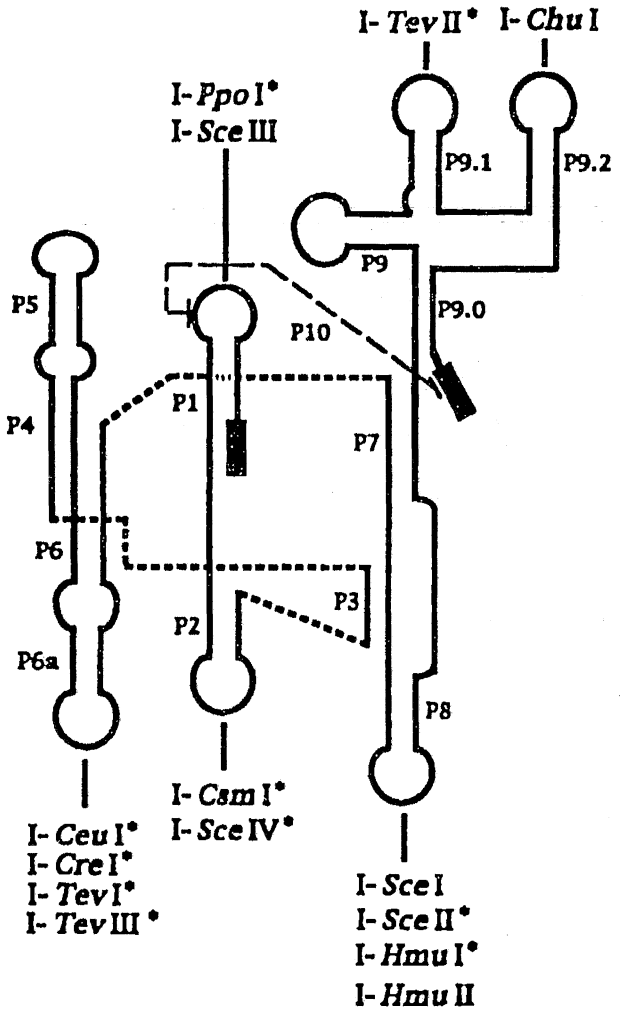

I-*Tev* II*
I-*Chu* I
I-*Ppo* I*
I-*Sce* III
P9.1
P9.2
P9
P9.0
P5
P10
P1
P4
P7
P6
P3
P2
P6a
P8
I-*Csm* I*
I-*Sce* IV*
I-*Ceu* I*
I-*Cre* I*
I-*Tev* I*
I-*Tev* III*
I-*Sce* I
I-*Sce* II*
I-*Hmu* I*
I-*Hmu* II

Figure 1.5 Amino acid motif at the intein-extein boundaries.

Homing Endonuclease	Extein	Intein	Extein
PI–*Sce*I	- C	- . . . . - VHN	C -
PI–*Tli*I	- S	- . . . . - VHN	T -
PI–*Tli*II	- S	- . . . . - VHN	S -
PI–*Mtu*I	- C	- . . . . - VHN	C -

Figure 1.6 Chemical rearrangements of the amino acid asparagine–a model for protein splicing. The cleavage pathway shown on the right may be involved in protein splicing.

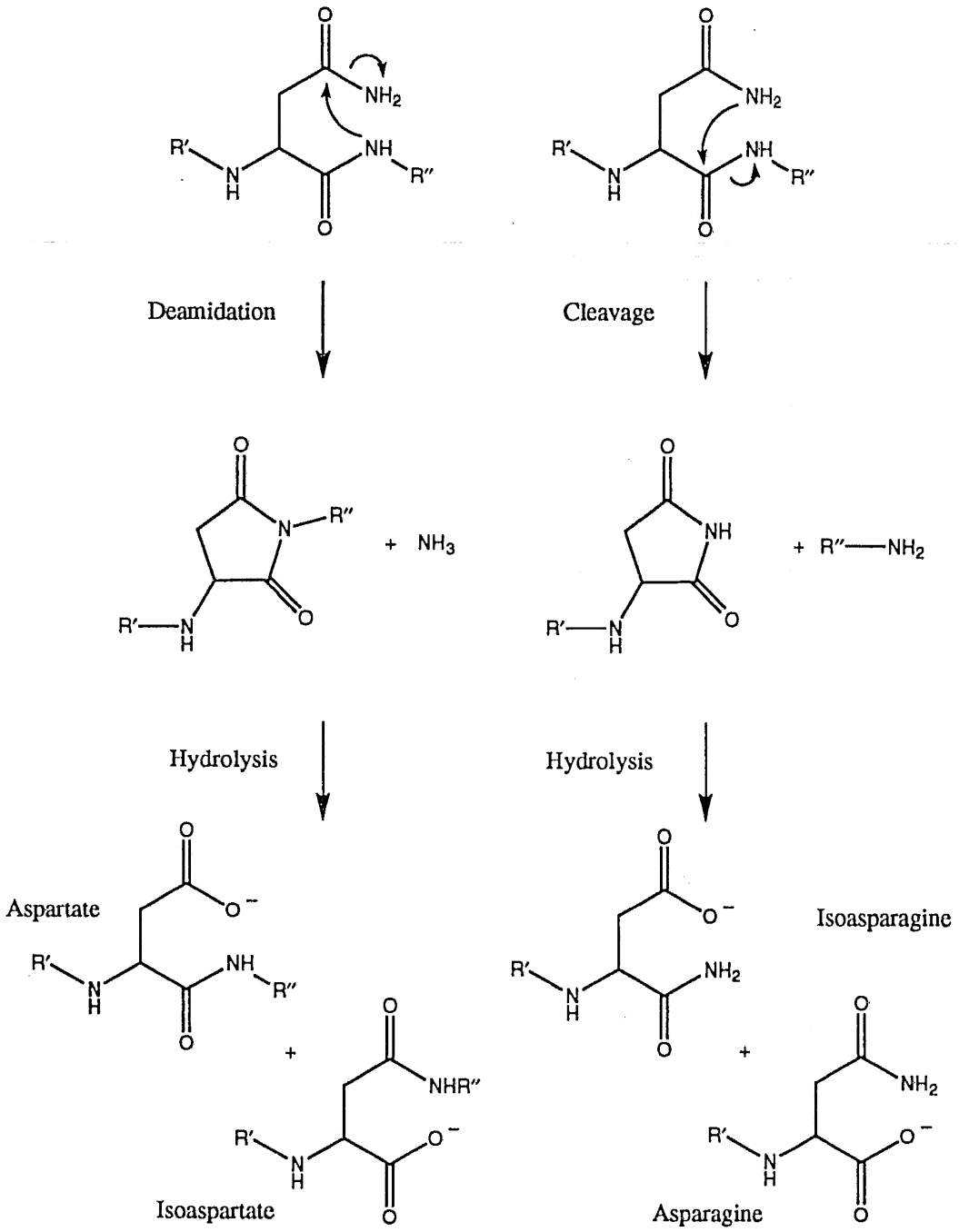

Deamidation
Cleavage
+ NH3
+ R″—NH2
Hydrolysis
Hydrolysis
Aspartate
Isoasparagine
Isoaspartate
Asparagine

Figure 1.7 Sequence Motifs of Homing Endonucleases: the LAGLI-DADG motif, the GIY-YIG motif, and the H-N-H motif. The numerals indicate the number of amino acids preceding, separating, or following in the polypeptide sequence. Residues consistent with the conserved motif are underlined. I-*Ppo*I does not contain any of these motifs. This compilation has been updated from Mueller *et. al.*, (1993) and Gimble & Stephens, (1995).

LAGLI-DADG MOTIF

```
                        LAGLIDADG                LAGLIDADG
I-SceI      35 - AGIGLILGDAYI -  90 - LAYWFMDDGGKW - 86
I-SceII     86 - WLAGLIDGDGYF -  94 - WFVGFFDADGTI - 112
I-SceIII    30 - YLAGLIEGDGSI - 120 - WLAILTDADGNF - 160
I-SceIV    101 - IMTGILLTDGWI -  98 - SLAHMIMCDGSF - 83
PI-SceI    208 - YLLGLWIGDGLS -  96 - FLAGLIDSDGYV - 126
I-CsmI      42 - IAVGLLLSDAHA -  91 - ALAYWIAGDGCW - 80
I-CreI      11 - YLAGFVDGDGSI - 140
I-ChuI       9 - DIFGSLLGDGNL -  99 - ALAYFYIDDGAL - 86
I-CeuI      57 - FLAGFLEGEASL - 149
I-DmoI      12 - YLLGLIIGSGGL -  84 - FIKGLYVAEGDK - 73
I-PorI       5 - YVKALSAFDGYV -  67 - LLAAAIDAEGNV - 78
PI-TliI    146 - ELVGLIVGDGNW -  86 - FLRGLFSADGTV - 134
PI-TliII   281 - KLLGYYVSEGYA -  83 - FLEAYFEGDGDI - 150
PI-MtuI    113 - RDDGYLIGDGRD -  88 - LLFGLFESDGWV - 215
PI-PspI    279 - KLLGYYVSEGSA -  83 - FLEGYFIGDGDV - 151
HO Endo    214 - WMLGLWLGDGTT -  99 - FLAGLIDSDGYV - 250
Endo.SceI  203 - TLSGLIEGDGYI - 105 - WLAGFTAADGSR - 145
```

GIY-YIG MOTIF

```
                GIY          YIG
I-TevI    1 - KSGIYQIKNTLNNKVYVGSA..KDFEKR - 218
SegA      8 - YNYTYVITNLVNNKIYYGTHSTDDLNDG - 185
```

H-N-H MOTIF

```
                        E HH                N         H   H
I-TevIII   16-DSDGRTDEIHHKDGNR  ENNDL DNLMCLSIQEHYDIH-217
I-HmuI     64-EGYEEGLVVDHKDGNK  DNNLS TNLRWVTQKINVENQ-74
Eco CoE8  163-VGGRRSFELHHDKPISQDGGVYDMDNLRITTPKRHIDIH-3
Eco CoE9  163-VGGRKVYELHHDKPISQGGEVYDMDNIRVTTPKRHIDIH-3
Eco CoE2  539-VGGRERFELHHDKPISQDGGVYDMNNIRVTTPKRHIDIH-3
Eco Mcr   206-CENC14LEVHHVIPLSSGGADTT DNCVALCPNCHRELH-21
Sce gpII  723-CQIC 5LEVHHVRTLNNA  15  MNR 4ICKTCHFKVH-9
```

Figure 1.8 Homing site of I-*Ppo*I. The DNA bases are numbered about the dyad of pseudosymmetry. The bases which contribute to the symmetry are in bold. The arrow denotes the location of the Pp LSU3 intron insertion, and the site cleaved is indicated by a line.

	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	
5'	**C**	**T**	C	T	**C**	**T**	**T**	**A**	**A**	**G**	G	T	**A**	**G**	C	3'
3'	**G**	**A**	G	A	**G**	**A**	**A**	**T**	**T**	**C**	C	A	**T**	**C**	G	5'

Intron Insertion Site

Chapter 2

Substrate Binding and Turnover by the Highly Specific I-*Ppo*I Endonuclease

Previously published as:

Wittmayer, P. K. and Raines, R. T. Substrate Binding and Turnover by the Highly Specific I-*Ppo*I Endonuclease. *Biochemistry* 1996, **35**, 1076-1083.

FOOTNOTES AND ABBREVIATIONS

1. Abbreviations: bp, base pairs; C.N., coordination number; nt, nucleotides.

2.1 Introduction

Cells use proteins to access and maintain the information encoded within their genome. Proteins can bind to double-stranded DNA with extraordinary specificity. After binding, enzymes can cleave or modify DNA with extreme efficiency. Understanding the molecular basis of such DNA recognition and manipulation is crucial for understanding many aspects of cellular function. Much of our current knowledge of protein – DNA interaction has been derived from investigations of prokaryotic transcriptional factors. Studies of these systems indicate that a primary determinant of specificity arises from direct contacts between the protein sidechains and the DNA backbone and bases (Steitz, 1990). Additional elements affecting DNA recognition include conformational changes of the DNA (Travers, 1991), conformational changes of the protein (Spolar & Record, 1994), and the presence of ions (Woodhead *et al.*, 1981; Record *et al.*, 1991; Xu & Schildkraut, 1991; Vipond *et al.*, 1995).

Recently, a novel family of enzymes was added to the repertoire of proteins that bind to DNA with high specificity. These enzymes are the intron-encoded endonucleases. Similar to the more familiar bacterial restriction endonucleases, the intron-encoded endonucleases catalyze the sequence-specific cleavage of double-stranded DNA. The large size of their recognition sites, however, is more analogous to that of many transcriptional factors. Intron-encoded endonucleases have recognition sites that range from 15 to 39 base pairs (Perlman & Butow, 1989). This new class of enzymes provides an opportunity to examine both complex sequence recognition and catalysis in a single system.

Intron-encoded endonucleases have their coding sequences located either wholly or partially within a mobile group I intron. Their function is to initiate the homing of their encoding intron. [For a review, see: Dujon (1989).] Briefly, intron homing is the site-specific, replicative transfer of mobile introns from an intron-containing allele to a

cognate, intron-lacking allele. The endonuclease initiates this genetic event by catalyzing site-specific double-stranded cleavage in the intron-lacking allele. The sources of homing introns and their encoded endonucleases are numerous and diverse, having been isolated from various eukaryotic (Colleaux *et al.*, 1986; Delahodde *et al.*, 1989; Gauthier *et al.*, 1991; Sargueil *et al.*, 1991; Thompson *et al.*, 1992; Perea *et al.*, 1993; Schapira *et al.*, 1993), prokaryotic (Bell-Pedersen *et al.*, 1989), and archaeal organisms (Hodges *et al.*, 1992; Dalgaard *et al.*, 1993; Dalgaard *et al.*, 1994).

The intron-encoded endonuclease that is the focus of this study, I-*Ppo*I, was first isolated by Vogt and coworkers from the slime mold *Physarum polycephalum.* (Muscarella *et al.*, 1990). [For nomenclature, see: (Dujon *et al.*, 1989).] I-*Ppo*I is encoded by the third intron located within the extrachromasomal nuclear DNA encoding the large ribosomal RNA subunit. Like the other intron-encoded endonucleases, I-*Ppo*I recognizes a large asymmetric sequence. In the presence of Mg^{2+}, I-*Ppo*I cleaves within the DNA sequence CTCTCTTAA/GGTAGC, generating fragments with 4-nt, 3′-OH extensions (Muscarella *et al.*, 1990; Ellison & Vogt, 1993). The size of I-*Ppo*I is small relative to its 15-bp recognition site. The protein forms a 2 x 20 kDa dimer in solution (Ellison & Vogt, 1993). For comparison, the restriction endonuclease *Eco*RI is a 2 x 31 kDa protein with a cognate site of only 6 base pairs (Modrich & Zabel, 1976; Greene *et al.*, 1981; Newman *et al.*, 1981). Thus, the residues in I-*Ppo*I have much greater information content than do those in *Eco*RI or other type II restriction endonucleases.

In this study, we examine the conformational changes of DNA upon binding, and the effects of monovalent and divalent ions on DNA binding and turnover by the I-*Ppo*I endonuclease. We find that the I-*Ppo*I bends its substrate upon binding. The strength of the binding interaction is affected dramatically by NaCl, and this effect allows an estimation of the number of Coulombic interactions between I-*Ppo*I and DNA. In addition, we have determined the steady-state kinetic parameters for the I-*Ppo*I-catalyzed

cleavage of its recognition site and begun to examine the role of the requisite divalent cation in affecting catalysis of DNA cleavage.

2.2 Materials and Methods

Materials. I-*Ppo*I, purified from an *Escherichia coli* expression system, was a generous gift from Promega Corp. (Madison, WI). The endonuclease is the larger (185 amino acid residues) form of the protein produced in *Physarum polycephalum.* Concentrations of the endonuclease were determined by using the method of Bradford (1976)(Bradford, 1976) with a kit from BioRad (Richmond, CA) and bovine serum albumin as the standard. Stock solutions of endonuclease were stored in 25 mM Tris-HCl buffer, pH 7.4, containing NaCl (250 mM), EDTA (10 mM), DTT (1 mM), and glycerol (10% v/v), and were diluted into solutions of bovine serum albumin (200 μg/ml) and glycerol (10% v/v) prior to use.

Poly(dIdC)·poly(dIdC) was from Sigma Chemical (St. Louis, MO), and was used without further purification. [γ-^{32}P]ATP, 6000 Ci/mmol, was from NEN Research Products (Wilmington, DE). Sephadex G-50 NICK columns and fluorescein phosphoramidite were from Pharmacia (Uppsala, Sweden). Fluorescein antibody columns were made by Dr. John Hill (University of Wisconsin – Madison) using antibodies from Dr. Edward Voss (University of Illinois at Urbana). DNA was purified with a Mermaid kit from Bio100 (La Jolla, CA) after electrophoresis in agarose gels. T4 polynucleotide kinase and the restriction endonucleases were from either Promega (Madison, WI) or New England Biolabs (Beverly, MA). Shrimp alkaline phosphatase came from United States Biochemical (Cleveland, OH).

General methods. DNA oligonucleotides were synthesized with β-cyanoethyl phosphoramidite chemistry (Sinha *et al.*, 1984) on an Applied Biosystems Model 392 DNA-RNA synthesizer. Complementary oligonucleotides were annealed by heating them

to 90 °C in a solution of 10 mM Tris-HCl buffer, pH 8.0, containing NaCl (200 mM), and then allowing them to cool slowly to room temperature. For binding and turnover assays, duplex DNA was purified from native polyacrylamide (15% w/v) gels. The band corresponding to double-stranded DNA was excised and soaked overnight in 0.30 M sodium acetate buffer, pH 5.0, containing EDTA (1 mM). After soaking, the supernatant was extracted with phenol:chloroform (1:1 v/v) and precipitated with ethanol. DNA substrates were radiolabeled on their 5′-OH group by treatment with [γ-^{32}P]ATP and T4 polynucleotide kinase. Salts and unincorporated nucleotides were removed by chromatography on Sephadex G50 columns.

Gel mobility shift assay for DNA binding. Gel mobility shift assays were based on the retardation of the electrophoretic mobility of a ^{32}P-labeled DNA molecule upon the binding of a protein (Fried & Crothers, 1981; Garner & Revzin, 1981). These assays were performed essentially as described (Ausubel *et al.*, 1989). The DNA was a 42-bp duplex (**I**) with the sequence indicated in Figure 1. [^{32}P]**I** (20 pM) was incubated at 25 °C with increasing amounts of I-*Ppo*I in a solution of 10 mM Tris-HCl buffer, pH 7.5, containing NaCl (50 mM), EDTA (1 mM), glycerol (10% v/v), and bovine serum albumin (50 μg/ml). Mg^{2+} was omitted to prevent DNA cleavage. After 1 h, the binding reactions were loaded directly onto a polyacrylamide (4% w/v) gel running at 30 V/cm. The gels were dried and counted using a Betascope 630 Blot Analyzer from Betagen (Waltham, MA). The disappearance of the band corresponding to free DNA was followed to determine the equilibrium dissociation constant. The data were fit with the program SIGMAPLOT 4.16 to eq 2.1, which describes binding to a single site.

$$B = \frac{B_{\max} F}{K_d + F} \tag{2.1}$$

In eq 2.1, B is the concentration of bound protein, B_{max} is the total concentration of binding sites, F is the concentration of free protein, and K_d is the dissociation constant for specific binding (Attie & Raines, 1995). Dissociation constants were calculated for the endonuclease dimer.

Fluorescence polarization assay for DNA binding. Fluorescence polarization assays were based on the change of the rotational correlation time of a fluorophore-labeled DNA molecule upon the binding of a protein (LeTilly & Royer, 1993). The protein•DNA complex, due to its increased size, tumbles more slowly than does free DNA. The ensuing reduction in the rotational correlation time of the fluorophore causes an increase in polarization, which allows the binding to be monitored (LeTilly & Royer, 1993). As shown in eq 2.2, polarization (P) is defined as the ratio of the difference between the vertical ($I_{\parallel}$) and horizontal ($I_{\perp}$) emission components to their sum.

$$P = 10^3 \times mP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{2.2}$$

Both specific and nonspecific DNA substrates were used in the fluorescence polarization assays: a 42-bp binding-site oligonucleotide (**I**; Figure 1), a self-complementary 60-nt binding-site oligonucleotide (**II**), and a 25-bp dIdC oligonucleotide [$(dIdC)_{25}\cdot(dIdC)_{25}$]. The fluorescein tag was linked to oligonucleotides by a six-carbon spacer to the terminal 5′-OH group of one strand of each duplex. Fluorescein was incorporated from its phosphoramidite in the final coupling step of DNA synthesis. To increase the fluorescein specific activity, the fluorescein-labeled strand was purified by chromatography on an immobilized anti-fluorescein antibody column. Briefly, the strand was loaded onto the column in 10 mM sodium phosphate buffer, pH 6.0, containing EDTA (0.10 mM) and NaCl (0.10 M). The loaded column was washed with the same buffer, and the strand was then eluted by denaturing the antibodies with aqueous HCl (0.2 M). The resulting

solution was neutralized by adding KOH (0.2 M). The purified strand was annealed to its complement, and the resulting duplex was purified by native gel electrophoresis.

Fluorescence polarization was measured at room temperature [(25 ± 2) °C] on a Beacon™ Fluorescence Polarization System (PanVera; Madison, WI) with excitation at 488 nM and emission at 520 nM. Fluorescein-labeled DNA duplex (450 – 550 pM) was incubated in a solution (1.10 mL) of 25 mM BisTris/Tris/BisTris Propane-HCl buffer, pH 7.5, containing EDTA (2 mM), DTT (1 mM), and the indicated concentration of NaCl. (This buffer system was used to allow for future analyses of pH-dependence.) Aliquots (2 – 5 μL) of I-*Ppo*I were added successively to the solution, which was then allowed to equilibrate. In each assay, the I-*Ppo*I concentration varied from 10 pM to 5 μM but the total volume of the solution varied by less than 10%. Three to six polarization measurements were made at each protein concentration. The dissociation constants were determined by fitting the data to eq 2.1 or eq 2.3, which describes binding to a single specific site as well as non-specific sites:

$$B = \frac{B_{max} F}{K_d + F} + K_{ns} F \qquad (2.3)$$

In eq 2.3, K_{ns} is a constant for nonspecific binding (Attie & Raines, 1995).

Circular permutation assay for DNA bending. The plasmid, pBend3 and the poly(dA)·poly(dT)-tract standards were gifts from Dr. Rick Gourse (University of Wisconsin – Madison). pBend3 contains the backbone of pBluescript SK- with the tandem repeat region of pBend2 (Kim *et al.*, 1989) cloned into the *Eco*RI and *Hind*III sites. To insert the recognition site of I-*Ppo*I, synthetic oligonucleotides containing the 42-bp sequence of I were cloned into pBend3 that had been digested with *Sal*I/*Xba*I. Five different 162-bp fragments were generated from the resulting plasmid using the *Mlu*I, *Nhe*I, *Eco*RV, *Nru*I, and *Bam*HI restriction endonucleases. In these isomeric fragments,

the I-*Ppo*I recognition site is inserted 126, 114, 78, 54, and, 23 bp from the end, respectively. The 5′-phosphoryl group of each fragment was hydrolyzed by treatment with alkaline phosphatase. The dephosphorylated fragments were purified from an agarose gel, and then labeled with ^{32}P.

Each ^{32}P-labeled fragment (5000 cpm; <500 pg) was incubated for 30 min at room temperature in a solution (50 μL) of 10 mM Tris-HCl buffer, pH 7.5, containing I-*Ppo*I (50 nM), poly(dIdC)·poly(dIdC) (0.2 μg/mL), EDTA (1 mM), NaCl (50 mM), bovine serum albumin (50 μg/mL), and glycerol (10% v/v). The samples were then loaded directly onto a native TBE polyacrylamide (8 or 10% w/v) gel. After electrophoresis, the gels were dried and visualized by using autoradiography.

A standard curve of DNA bending was obtained by using poly(dA)·poly(dT)-tract standards. Poly(dA)·poly(dT)-tracts, which each induce an 18° bend from linearity (Nelson *et al.*, 1987; Zahn & Blattner, 1987; Koo & Crothers, 1988; Koo *et al.*, 1990), were phased at 10-bp intervals such that the bends were cumulative. Fragments containing two, three, four, and five poly(dA)·poly(dT)-tracts were prepared from plasmids pJT170-2, pJT170-3, pJT170-4, and pJT170-5, respectively, as described previously (Thompson & Landy, 1988). Digesting each plasmid with *Nde*I or *Bam*HI created two fragments, one with the poly(dA)·poly(dT)-tract distortions near the end of the fragments and one with the distortions near the center, respectively. The two fragments from each plasmid were combined and subjected to electrophoresis alongside the I-*Ppo*I•DNA complex in a polyacrylamide (8% w/v) gel.

Steady-state kinetic analysis of catalysis. The ability of I-*Ppo*I to catalyze the cleavage of its recognition site was assayed by monitoring the cleavage of ^{32}P-labeled I. Reactions were performed at 37 °C in solutions (100 μL) of 10 mM Tris-HCl buffer, pH 7.5, containing I-*Ppo*I (5 – 100 pM), [^{32}P]I (10 pM – 50 nM), $MgCl_2$ (10 mM), DTT (1 mM), NaCl (0, 50, 100, or 200 mM), and bovine serum albumin (50 μg/mL). The

substrate concentration was at least 10-fold greater than the enzyme concentration in each reaction, except for that of the lowest substrate concentration in the absence of NaCl. (Accordingly, only a lower limit could be determined for the value of K_m in the absence of NaCl.) Aliquots (10 μL) were quenched at various times (5 min – 4 h) by addition to a solution (10 μL) of SDS (2% w/v), EDTA (100 mM), glycerol (20% v/v), and bromophenol blue (0.01% w/v). Products were separated from reactants on a polyacrylamide (15% w/v) gel and quantified as in the gel mobility shift assays. Initial velocities were measured at ≥6 different substrate concentrations at each NaCl concentration. To obtain steady-state kinetic parameters, the initial velocity data were fit to the Michaelis–Menten equation with the program HYPERO (Cleland, 1979). Parameters were calculated for the endonuclease dimer.

Assays for metal-ion dependence of catalysis. A variety of divalent cations were examined for their ability to substitute for Mg^{2+} and permit the sequence-specific cleavage of DNA by I-*Ppo*I. The reactions were performed in 10 mM Tris-HCl buffer, pH 7.5, containing [^{32}P]I (10 nM), I-*Ppo*I (2 nM), and the chloride salt of Mg^{2+}, Mn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}, Ca^{2+}, or Zn^{2+} (10 mM of the puratronic grade from Johnson Mathey; Ward Hill, MA). The reactions were incubated at 37 °C and quenched after 1 h. In all manipulations, care was taken to avoid contaminating metal ions: glassware was soaked overnight in 20% (v/v) nitric acid and rinsed thoroughly with 18-MΩ·cm water; plasticware was soaked in a solution of EDTA (10 mM) overnight, and rinsed thoroughly with 18-MΩ·cm water; and all buffers were passed through a column (10 mL) of Chelex-100 resin (Bio-Rad; Richmond, CA).

2.3 Results

*DNA binding by I-*Ppo*I.* In the absence of Mg^{2+}, I-*Ppo*I binds to DNA but is unable to catalyze cleavage. The formation of an I-*Ppo*I•DNA complex can therefore be

followed by eliminating Mg^{2+} from the reaction buffer. Two independent techniques were used to monitor DNA binding: gel mobility shift assays (Fried & Crothers, 1981; Garner & Revzin, 1981) and fluorescence polarization assays (LeTilly & Royer, 1993).

Gel mobility shift assays established that I-*Ppo*I retained specificity for its recognition site in the absence of Mg^{2+}. As shown in Figure 2A, an I-*Ppo*I•DNA complex was formed at low I-*Ppo*I concentrations. Unlabeled **I**, but not poly(dIdC)·poly(dIdC) (0.2 μg/ml), competed with [^{32}P]**I** for complex formation at low concentrations of I-*Ppo*I (data not shown). This complex is therefore a specific complex, and it has $K_d = 0.7$ nM. At high I-*Ppo*I concentrations, the I-*Ppo*I•DNA band was smeared, apparently due to the formation of less-stable, higher-order complexes. These nonspecific complexes dissociated upon addition of poly(dIdC)·poly(dIdC) (0.2 μg/ml) (data not shown).

Both nonspecific and specific DNA-binding by I-*Ppo*I were examined with fluorescence polarization assays. Three DNA fragments were used: **I**, **II**, and $(dIdC)_{25}\cdot(dIdC)_{25}$. The fluorescein and linker in **I** and **II** did not interfere with binding by I-*Ppo*I as demonstrated by gel mobility shift assays (data not shown). The binding of **I** and $(dIdC)_{25}\cdot(dIdC)_{25}$ in the presence of low ionic strength (33 mM added NaCl) is shown in Figure 2B. A small increase in polarization (10 mP) was observed upon binding to the specific recognition site of **I**. [The results from **II** (data not shown) were similar to those from **I**, except that the change in polarization upon binding to **II** was approximately 60% larger.] This initial increase was followed by a second increase in polarization from nonspecific binding. The polarization did not reach a plateau at the highest protein concentration examined (1 μM), which indicates that the nonspecific sites were not saturated. The polarization was also observed to increase upon binding to $(dIdC)_{25}\cdot(dIdC)_{25}$. Nonspecific binding of I-*Ppo*I to **I** and $(dIdC)_{25}\cdot(dIdC)_{25}$ began at the same concentration of protein (50 nM). From the data in Figure 2, it is apparent that the

change in fluorescence polarization of a fluorescein-labeled duplex provides an effective measure of the binding of I-*Ppo*I to DNA in homogeneous solution.

The effect of increasing NaCl concentration on the binding of I-*Ppo*I to its recognition site was examined by using fluorescence polarization to monitor complex formation. As shown in Figure 3, at high salt concentrations, specific binding is weakened slightly but nonspecific binding is impaired drastically. Dissociation constants were determined for binding to the specific site, and these results are reported in Table 1. Only lower limits are reported for nonspecific dissociation constants due to the inability to saturate the nonspecific sites, even at high protein concentrations.

*DNA bending upon binding by I-*Ppo*I.* A circular permutation assay (Wu & Crothers, 1984) was used to detect protein-induced DNA bending by I-*Ppo*I. This assay is based on position-dependent effects of DNA distortions on the electrophoretic mobilities of DNA fragments. A fragment with a distortion located near the end migrates faster during electrophoresis than does an isomeric fragment with a distortion near the center. As shown in Figure 4A, the mobility of an I-*Ppo*I•DNA complex was dependent on the distance of the recognition site from the end of the fragment. The *Eco*RV fragment, which has the recognition site positioned in the center, migrated slowest when bound by I-*Ppo*I. The free DNA fragments did not display any variable mobility. The distortion of the recognition site is therefore not inherent to the DNA sequence itself but is induced upon binding of I-*Ppo*I.

The degree of the bend in DNA induced by I-*Ppo*I was determined from eq 2.4 (Thompson & Landy, 1988).

$$\mu_{rel} = \cos\left(\frac{\alpha}{2}\right) \tag{2.4}$$

In eq 2.4, α is the degree to which linear DNA ($\alpha = 0$) is bent. The relative mobilities of the complexes in which I-*Ppo*I was bound to the *Mlu*I or *Bam*HI fragments (μ_{end}) and the *Eco*RV fragment (μ_{middle}) were used to calculate $\mu_{rel} = \mu_{middle}/\mu_{end} = 0.946 \pm 0.011$, where the standard deviation was based on eight independent experiments. Substituting this value for μ_{rel} into eq 2.4 reveals that I-*Ppo*I bends DNA by $\alpha = 38° \pm 4°$ upon binding to its recognition site. To verify this value, the relative mobility of the I-*Ppo*I•DNA complex was compared to that of poly(dA)·poly(dT)-tract standards. As shown in Figure 4B, the relative mobilities of the five fragments containing the poly(dA)·poly(dT)-tracts were within error of that predicted by eq 2.4. Thus, eq 2.4 accurately predicts the degree of bending for our system.

*Steady-state kinetic parameters for catalysis by I-*Ppo*I.* In the presence of Mg^{2+}, I-*Ppo*I catalyzes the hydrolysis of two phosphodiester bonds within its recognition site, generating 4-nt, 3′-OH extensions. The steady-state kinetic parameters for this reaction are reported in Table 2. The highest value of k_{cat}/K_m was observed at low concentrations of NaCl. As the NaCl concentration increased to 200 mM, the value of k_{cat}/K_m decreased by at least 200-fold. The decrease in k_{cat}/K_m (from >4.5 $nM^{-1}min^{-1}$ to 0.023 $nM^{-1}min^{-1}$) was a result of a 400-fold increase in K_m and a twofold increase in k_{cat}. These results suggest that NaCl influences the binding rather than the turnover of substrate.

*Metal-ion-dependence of catalysis by I-*Ppo*I.* A divalent cation, often Mg^{2+}, is required for most reactions catalyzed by DNA-modifying enzymes (Roberts & Halford, 1993). The specificity of the enzymes for the cofactor is, however, variable. Seven divalent cations, including Mg^{2+}, were tested for their ability to act as cofactors in the I-*Ppo*I-catalyzed reaction. All cations tested except for Cu^{2+} supported catalysis with various success. The order of activity was $Mg^{2+} > Mn^{2+} > Ca^{2+} = Co^{2+} > Ni^{2+} > Zn^{2+}$. Catalysis with each of these cofactors gave the expected cleavage products, indicating

that substrate specificity was not entirely metal-ion-dependent (Figure 5). Thus, other divalent cations can substitute for Mg^{2+}, but with a decrease in catalytic efficiency.

2.4 Discussion

*Comparison of catalysis by I-*Ppo*I and restriction endonucleases.* Catalysis by both I-*Ppo*I (Table 2) and many restriction endonucleases depends greatly on reaction conditions. A comparison of steady-state kinetic parameters is therefore a challenge (Anderson *et al.*, 1981). Nonetheless, considering only the optimal conditions tested here for I-*Ppo*I (that is, low concentrations of NaCl at pH 7.5) and optimal conditions for the respective restriction endonucleases, the value of k_{cat}/K_m for I-*Ppo*I is similar of that of many restriction endonucleases. The value of k_{cat}/K_m for I-*Ppo*I is ≥ 4.5 $nM^{-1}min^{-1}$ or $\geq 7.5 \times 10^7$ $M^{-1}s^{-1}$ (Table 2). The values of k_{cat}/K_m for the cleavage of plasmid DNA by *Eco*RI, *Eco*RV, and *Bam*HI have been reported to be 5×10^7, 3×10^7, and 7×10^6 $M^{-1}s^{-1}$, respectively (Halford & Johnson, 1981; Nardone & Chirikjian, 1987; Halford & Goodall, 1988). These values of k_{cat}/K_m approach that expected for diffusion, which limits the value of all second-order rate constants (Blacklow *et al.*, 1988).

Although the value of k_{cat}/K_m for catalysis by I-*Ppo*I is similar to those for various restriction endonucleases, the values of k_{cat} and K_m differ. I-*Ppo*I has a lower Michaelis constant accompanied by a lower turnover number. Under the optimal conditions used here, I-*Ppo*I has a K_m of ≤ 0.01 nM. In contrast, *Eco*RI, *Eco*RV, and *Bam*HI have K_m's of 2.5, 0.5, and 3.6 nM, respectively. The higher affinity of I-*Ppo*I for its substrate could reflect the larger size of its recognition site. The value of k_{cat} for I-*Ppo*I is 0.046 min^{-1}, while the values of k_{cat} for *Eco*RI, *Eco*RV, and *Bam*HI, are 7.7, 0.9, 1.5 min^{-1}, respectively. Under other reaction conditions, however, the values of the Michaelis constant and turnover number for catalysis by I-*Ppo*I are similar to those for restriction endonucleases. For example, at 200 mM NaCl and pH 7.5, the K_m of I-*Ppo*I

increases to 4 nM (Table 2). At pH 10, the k_{cat} of I-*Ppo*I increases to 2.6 min^{-1} (Lowery *et al.*, 1992).

*I-*Ppo*I•DNA in the absence of* Mg^{2+}. The Mg^{2+} cofactor is not required for specific binding of I-*Ppo*I to its recognition site. In the absence of Mg^{2+}, I-*Ppo*I is unable to catalyze the cleavage of DNA. Still, the enzyme maintains enough contacts with DNA to favor its cognate sequence over noncognate sequences. Although many restriction endonucleases form stable complexes with DNA in the absence of Mg^{2+}, they do not necessarily bind specifically. For example, *Eco*RV, *Taq*I, *Cfr*9I, and *Bcg*I bind equally well to DNA containing or lacking their recognition sites (Taylor *et al.*, 1991; Zebala *et al.*, 1992; Siksnys & Pleckaityte, 1993; Kong *et al.*, 1994). In this respect, I-*Ppo*I is similar to the restriction endonucleases *Eco*RI, *Bam*HI, *Rsr*I, *Fok*I, and *Sma*I (Halford & Johnson, 1980; Terry *et al.*, 1983; Aiken *et al.*, 1991; Xu & Schildkraut, 1991; Waugh & Sauer, 1993; Withers & Dunbar, 1995), which bind to their recognition sites specifically in the absence of a divalent metal ion.

The topology of the DNA in the I-*Ppo*I•DNA complex formed in the absence of Mg^{2+} deviates from that of typical B-form DNA. I-*Ppo*I induces a bend of 38° ± 4° upon binding (Figure 4). The distortion of DNA observed here is consistent with the results of previous footprinting studies, which revealed two regions of increased sensitivity to methylation by dimethylsulfate, a small chemical probe (Ellison & Vogt, 1993). These hypersensitive regions are located on both sides [A(+5) and G(–3), A(–4)] of the central six base pairs, and may be the origins for the observed bend of the recognition site by I-*Ppo*I. The ease with which a DNA sequence adopts this requisite conformation may provide a level of discrimination for the enzyme between cognate and noncognate sites. In addition, this distortion may enable the small I-*Ppo*I protein to make contact with its large recognition site, which could span nearly 1.5 helical turns of DNA.

Some restriction endonucleases are also known to distort the double helix. For example, x-ray diffraction analysis of the *Eco*RI•DNA complex shows that this enzyme unwinds three base pairs, opening up the helical grooves and creating three separate kinks (Frederick *et al.*, 1984). *Eco*RV creates a central 50° kink and compresses the major groove of its cognate DNA, but does not distort noncognate DNA (Winkler *et al.*, 1993). In contrast, *Pvu*II and *Bam*HI do not distort their recognition sites (Cheng *et al.*, 1994; Newman *et al.*, 1995). The shape, direction, and groove adjustments of the bent I-*Ppo*I recognition site await direct structural analysis.

Gel mobility shift assays are valuable tools for visualizing individual species and extracting information on the binding and bending of DNA by I-*Ppo*I. These assays suffer, however, from an inability to mimic true solution conditions. To complement the binding data that are obtained from mobility shift assays, we used a second technique to measure binding.[1] Fluorescence polarization assays can be used to measure binding in solution without separating bound from free species. For I-*Ppo*I•DNA, the results obtained from the two independent methods are in gratifying agreement. In both assays, specific binding by I-*Ppo*I is observed when the concentration of endonuclease is <10 nM, and nonspecific binding is observed when the concentration is >50 nM (Figure 2). In the presence of 50 mM NaCl (which is the NaCl concentration in the incubation buffer used for the gel mobility shift assays), the dissociation constant measured with florescence polarization (K_d = 3.3 nM) is within 5-fold of that determined from the gel mobility shift assay (K_d = 0.7 nM).

*Effect of NaCl on I-*Ppo*I•DNA*. The affinity of I-*Ppo*I for its recognition site is influenced dramatically by the concentration of NaCl (Figure 3). This change in affinity is likely to be a result of the polyanionic nature of the DNA helix. The binding of DNA

---

1. Filter-binding assays were unsuccessful because the I-*Ppo*I•DNA complex was not retained by nitrocellulose filters.

by a protein is accompanied by the release of cations from the DNA. The accompanying increase in entropy is a driving force for binding (Record *et al.*, 1976). Theory predicts that a protein•nucleic acid complex will dissociate with increasing cation concentration as in eq 2.5.

$$\frac{\partial \log K_d}{\partial \log[\text{cation}]} = k + m' \psi \tag{2.5}$$

In eq 2.5, k is the number of anions displaced upon binding, m' is the number of cations displaced upon binding, and Ψ is a measure of counterion binding to the nucleic acid, which is 0.88 for double-helical native DNA (Record *et al.*, 1991). Our values of K_d and cation concentration from Table 1 are plotted in Figure 6. This plot shows two regions of linearity—a segment at higher salt concentrations for which the affinity of I-*Ppo*I declines with increasing salt and a segment at lower salt concentrations for which there is little change in affinity. Using the values of K_d at higher salt concentrations, $m' = 6.3 \pm 1.4$, which suggests that approximately 6 cations are displaced from DNA upon binding of I-*Ppo*I at pH 7.5.

At salt concentrations <150 mM, our data do not follow eq 2.5. Similar deviations have been observed for the binding of other proteins to DNA at low salt concentrations (Jen-Jacobson *et al.*, 1983; Murdoch *et al.*, 1991). These deviations may result from aggregation of the protein or from exceeding the detection limits of the technique used to measure binding. For example, at low ionic strength, the precipitation of *Eco*RI gives an artificially high value of K_d (Jen-Jacobson *et al.*, 1983). We have observed no evidence that I-*Ppo*I precipitates or aggregates in solutions of low ionic strength. Alternatively, the true dissociation constant of I-*Ppo*I at low salt concentrations may be beyond the detection limits of our assays. If virtually all added protein binds to DNA, then an accurate measure of dissociation constants is difficult to obtain because so little protein

remains free in solution. Our analysis of catalysis by I-*Ppo*I indicates that the affinity of I-*Ppo*I for DNA does indeed become even tighter at salt concentrations <150 mM (Table 2 and Figure 6).

*Effect of NaCl on catalysis by I-*Ppo*I.* Catalysis by restriction endonucleases is usually sensitive to ionic strength, and that by I-*Ppo*I is not an exception. The value of k_{cat}/K_m decreases by 200-fold upon addition of 200 mM NaCl. This decrease comes primarily from the diminished affinity of the enzyme for its recognition site—K_m decreases by 400-fold. Concomitantly, k_{cat} increases by twofold. Although the origin of this nominal increase is not clear, Na^+ may serve to facilitate product release, which has been postulated to be the rate-limiting step in catalysis by the intron-encoded endonuclease I-*Sce*I (Perrin *et al.*, 1993).

*Effect of metal ions on catalysis by I-*Ppo*I.* Although a divalent cation is required for catalysis by I-*Ppo*I, the enzyme does not have a strict preference for a particular metal ion. All metal ions tested, excluding Cu^{2+}, support catalysis (Figure 5). This result is in contrast to catalysis by *Eco*RI and *Eco*RV. The activities of these two enzymes are relatively metal-ion-dependent, and Ca^{2+} does not support catalysis despite binding to the same site as do other metal ions (Vipond *et al.*, 1995; Vipond & Halford, 1995).

The metal-ion binding site of I-*Ppo*I is plastic. The enzyme binds productively to divalent cations with radii as small as Ni^{2+} [0.69 Å, C.N. 6; (Shannon, 1974)] and as large as Ca^{2+} [1.00 Å, C.N. 6]. Preferences in the metal-ion binding site are, however, apparent. The cations best capable of supporting catalysis, Mg^{2+}, Mn^{2+}, Co^{2+}, and Ca^{2+}, are oxophilic and can adopt an octahedral coordination geometry. In contrast, the other divalent ions tested are thiophilic or prefer other coordination geometries. Thus, I-*Ppo*I may be similar to *Eco*RV, which binds Mg^{2+} in octahedral coordination with the oxygens from the scissile phosphoryl group, the sidechains of Asp74 and Asp90, and three water molecules as its ligands (Kostrewa & Winkler, 1995). Asp74 and Asp90 of *Eco*RV are

part of the active site motif: P(D/E)···(D/E)XK, which also appears in other type II restriction endonucleases (Thielking *et al.*, 1991; Anderson, 1993), and in I-*Ppo*I. The sequence alignment in Table 3 suggests that Asp109 and Asp140 of I-*Ppo*I may coordinate to Mg^{2+} or another divalent metal ion.

Conclusions. This study represents the first detailed biochemical analysis of binding and catalysis by an intron-encoded endonuclease. We have found that tight, specific binding by I-*Ppo*I induces a conformational change in DNA in the absence of Mg^{2+}. In addition, the interaction of I-*Ppo*I with DNA, both in equilibrium and during steady-state catalysis, weakens with increasing concentrations of monovalent cations in a manner consistent with the entropy gained from the release of cations from DNA being a driving force for binding. Catalysis by I-*Ppo*I is efficient, with values of k_{cat}/K_m at low concentrations of NaCl approaching the diffusion limit. Finally, the metal-ion binding site of I-*Ppo*I prefers oxophilic, octahedral divalent cations, but is otherwise nonspecific. Our work demonstrates that the recent discovery of intron-encoded endonucleases such as I-*Ppo*I provides a unique opportunity to dissect both enzymatic catalysis and protein – DNA interactions in a *single* system.

Table 2.1: Effect of NaCl on Binding of DNA by I-*Ppo*I [a]

[NaCl] (mM)	$K_d^{specific}$ (nM)	$K_d^{nonspecific}$ (nM)
33	5.4 ± 2.4	>175
50	3.3 ± 2.5	>350
100	3.4 ± 2.6	>550
150	3.7 ± 1.7	>1000
175	5.4 ± 3.1	>1000
200	36.0 ± 9.6	>1000
225	20.1 ±11.0	>1000
250	111.7± 31.4	>1000
275	69.3 ± 19.0	>1000

[a] Determined by fluorescence polarization in 25 mM Bis-Tris/Tris/Bis-Tris propane buffer, pH 7.5, containing EDTA (2 mM), DTT (1 mM), BSA (50 μg/mL), and NaCl (as indicated).

Table 2.2: Effect of NaCl on Steady-State Kinetic Parameters for DNA Cleavage by I-*Ppo*I [a]

[NaCl] (mM)	k_{cat} (min^{-1})	K_m (nM)	k_{cat}/K_m ($nM^{-1}min^{-1}$)
0	0.046 ± 0.003	≤0.01	≥4.5
50	0.060 ± 0.010	0.05 ± 0.02	1.2 ± 0.4
100	0.065 ± 0.005	0.3 ± 0.1	0.22 ± 0.03
200	0.090 ± 0.005	4.0 ± 0.7	0.023 ± 0.002

[a] Determined by gel electrophoresis in 10 mM Tris-HCl buffer, pH 7.5, containing $MgCl_2$ (10 mM), DTT (1 mM), BSA (50 μg/mL), and NaCl (as indicated).

Table 2.3: Sequence Alignment of I-*Ppo*I with Type II Restriction Endonucleases

endonuclease	putative active-site motif [a]		
I-*Ppo*I	$P_{108}D_{109}$	···	$D_{140}N_{141}K_{142}$
*Eco*RI	$P_{90}D_{91}$	···	$E_{111}A_{112}K_{113}$
*Eco*RV	$P_{73}D_{74}$	···	$D_{90}I_{91}K_{92}$
*Eco*571	$P_{77}D_{78}$	···	$D_{92}A_{93}K_{94}$
	$P_{811}D_{812}$	···	$D_{833}Q_{834}K_{835}$
*Fok*I	$P_{449}D_{450}$	···	$D_{467}T_{468}K_{469}$
*Hha*II	$P_{98}D_{99}$	···	$E_{116}L_{117}K_{118}$
*Kpn*I	$P_{147}D_{148}$	···	$D_{163}P_{164}K_{165}$
*Ngo*PII	$P_{72}D_{73}$	···	$E_{88}S_{89}K_{90}$
*Pst*I	$P_{304}D_{305}$	···	$D_{320}K_{321}K_{322}$
*Rsr*I	$P_{95}D_{96}$	···	$E_{116}S_{117}K_{118}$

[a] Based on Thielking *et al.* (1991) and Anderson (1993).

Figure 2.1 DNA used for assays of binding and catalysis by I-*Ppo*I. (Top) **I**, a 42-bp duplex. (Bottom) **II**, a self-complementary 60-nt strand. The site recognized by I-*Ppo*I is numbered –7 to +8, and the site cleaved is indicated by a line. F indicates the location of the fluorescein tag used for the fluorescence polarization assays. No fluorescein is present in other assays.

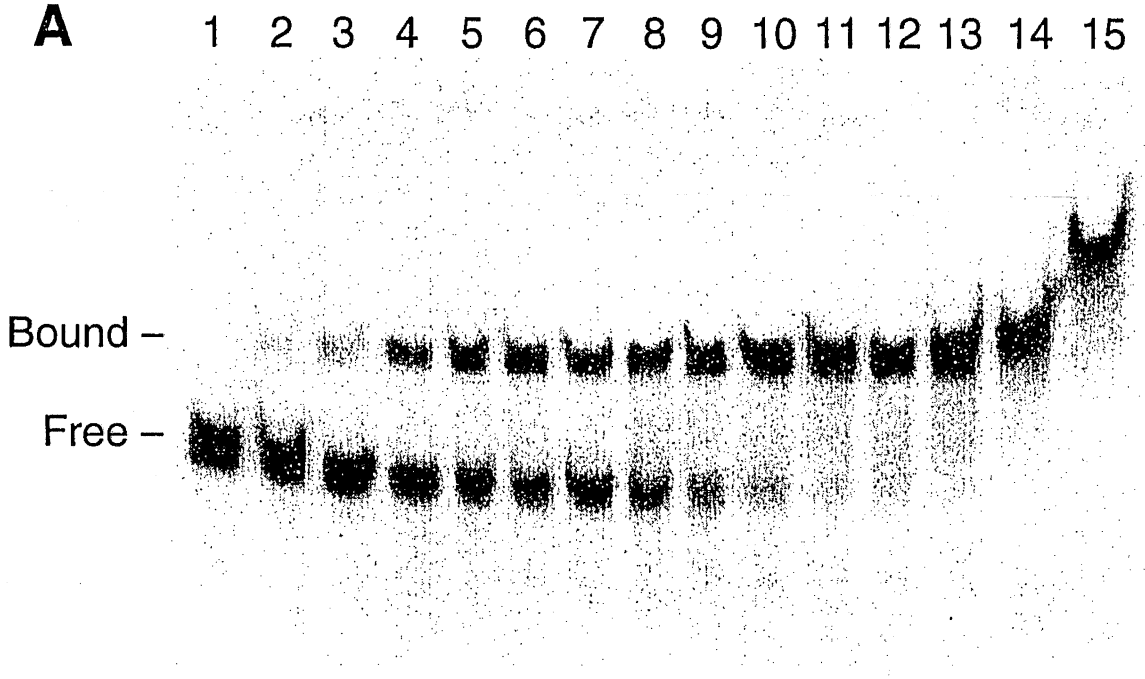
```
   F
   |                  -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8
5' AATTAGTAACTATGA   C  T  C  T  C  T  T  A  A  G  G  T  A  G  C   CAAATGCCTGCA 3'
3' TTAATCATTGATACT   G  A  G  A  G  A  A  T  T  C  C  A  T  C  G   GTTTACGGACGT 5'
```

I

```
      F
      |
   5' GCATGA   C  T  C  T  C  T  T  A  A  G  G  T  A  G  C   CAAATGC^T T
   3' CGTACT   G  A  G  A  G  A  A  T  T  C  C  A  T  C  G   GTTTACG_T T
```

II

Figure 2.2 Two assays of the binding of I-*Ppo*I to **I**. (A) Gel mobility shift assay of binding of I-*Ppo*I to [^{32}P]**I** in 10 mM Tris-HCl buffer, pH 7.5, containing NaCl (50 mM), EDTA (1 mM), glycerol (10% v/v), and bovine serum albumin (50 μg/ml). The lanes, 1 – 15, have increasing concentrations of dimeric I-*Ppo*I: 0, 0.02,.0.05, 0.08, 0.13, 0.20, 0.32, 0.50, 0.79, 1.3, 2.0, 3.2, 5.0, 50, and 500 nM. The apparent K_d for specific binding is 0.7 nM. (B) Fluorescence polarization assay of binding of I-*Ppo*I to fluorescein-labeled **I** (●) or fluorescein-labeled $(dIdC)_{25}\cdot(dIdC)_{25}$ (O). I-*Ppo*I was added to 25 mM BisTris/Tris/BisTris Propane-HCl buffer, pH 7.5, containing fluorescein-labeled DNA (450 – 550 pM), NaCl (33 mM), EDTA (2 mM), and DTT (1 mM). Each data point is an average of 3 – 6 measurements. The apparent K_d for the specific site is 5 nM.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Bound –
Free –

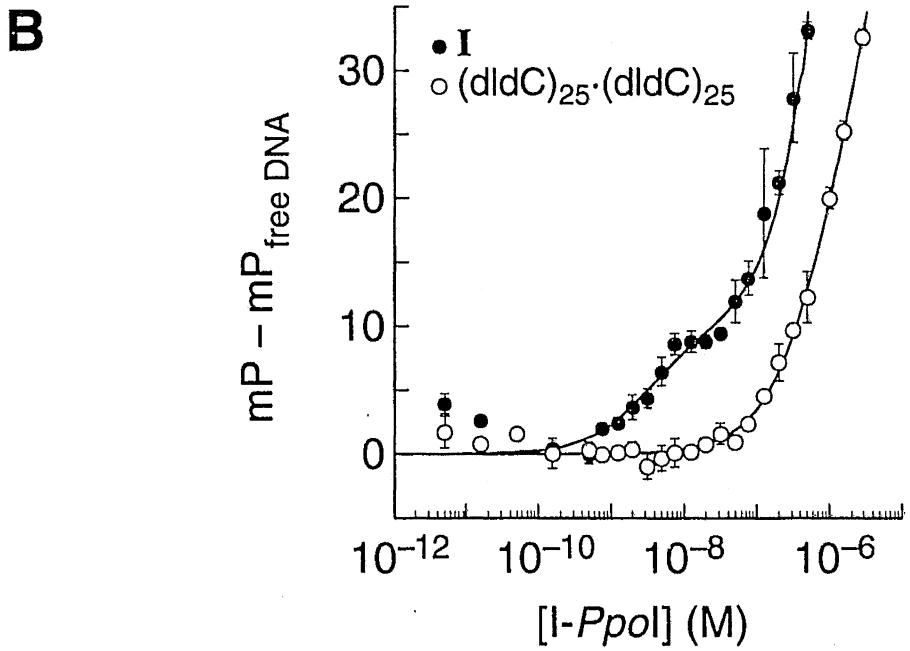

B
• I
○ (dIdC)25·(dIdC)25
mP – mPfree DNA
30
20
10
0
10−12
10−10
10−8
10−6
[I-Ppo I] (M)

Figure 2.3 Binding of I-*Ppo*I to **I** at three NaCl concentrations. I-*Ppo*I was added to 25 mM BisTris/Tris/BisTris Propane-HCl buffer, pH 7.5, containing fluorescein-labeled **I** (450 – 550 pM), NaCl [33 mM (●; as in Figure 2), 100 mM (O), or 275 mM (■)], EDTA (2 mM), and DTT (1 mM). Each data point is an average of 3 – 6 measurements. The lines are best fits to either eq 1 (275 mM) or eq 3 (33 and 100 mM).

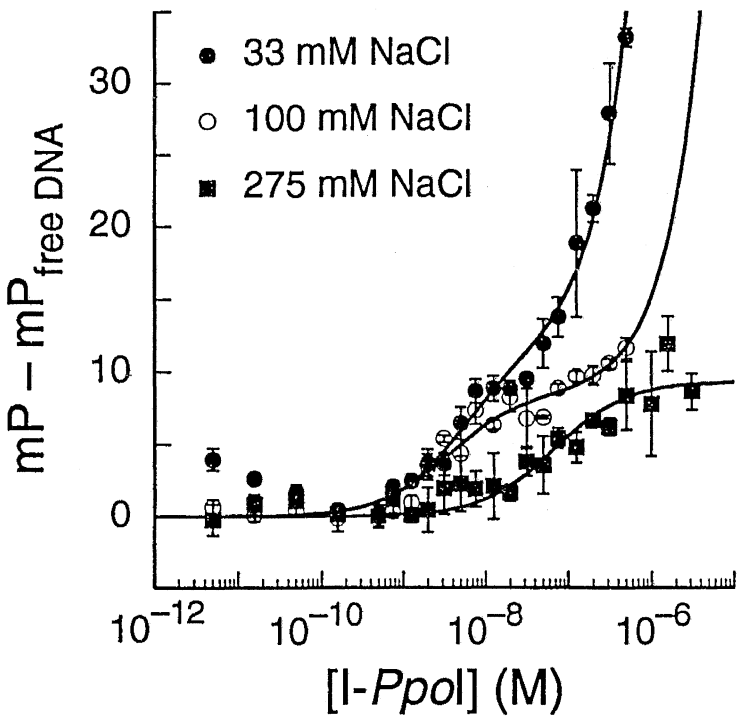

33 mM NaCl
100 mM NaCl
275 mM NaCl
mP – mP free DNA
30
20
10
0
10−12
10−10
10−8
10−6
[I-Ppol] (M)

Figure 2.4 Bending of DNA upon binding of I-*Ppo*I. (A) The first five lanes contain circularly-permuted DNA fragments generated with the restriction endonucleases: *Mlu*I, *Nhe*I, *Eco*RV, *Nru*I, and *Bam*HI. The last lane contains the fragments from both *Mlu*I and *Eco*RV. The I-*Ppo*I recognition site is located near the end and middle of the *Mlu*I and *Eco*RV fragments, respectively. (B) Relative mobilities of the *Mlu*I and *Eco*RV fragments bound to I-*Ppo*I (O) and fragments containing different numbers of poly(dA)·poly(dT)-tracts (●). Eq 4 was used to draw the curve and to calculate that I-*Ppo*I bends DNA by $\alpha = 38^\circ \pm 4^\circ$ upon binding to its recognition site.

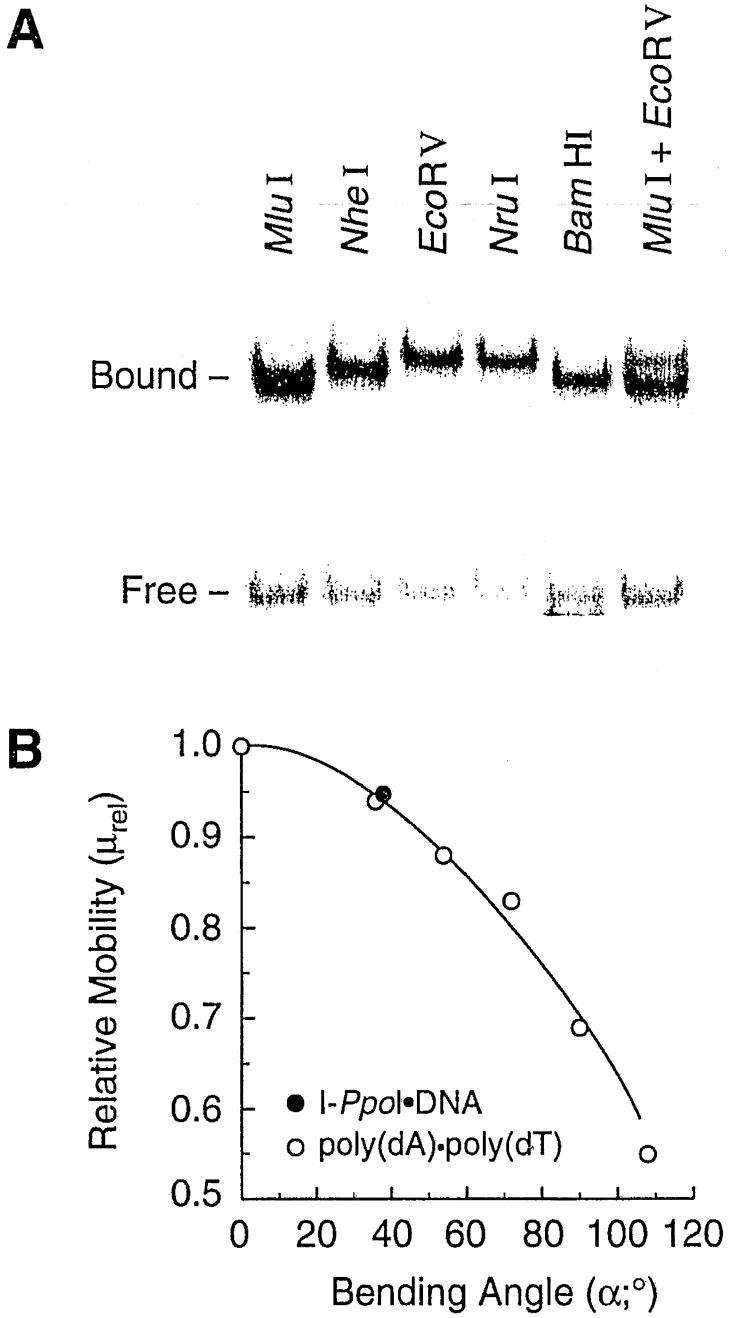

A
Mlu I
Nhe I
EcoR V
Nru I
Bam HI
Mlu I + EcoR V
Bound –
Free –
B
Relative Mobility (μ_{rel})
1.0
0.9
0.8
0.7
0.6
0.5
0
20
40
60
80
100
120
Bending Angle (α;°)
● I-Ppo I•DNA
○ poly(dA)•poly(dT)

Figure 2.5 Cleavage of **I** by I-*Ppo*I in the presence of various divalent cations. Percent cleavage is represented relative to the product formation in the reaction containing the Mg^{2+} ion. The reaction containing the Mg2+ ion represents 100% cleavage. Each reaction yielded products of 22 and 20 bp.

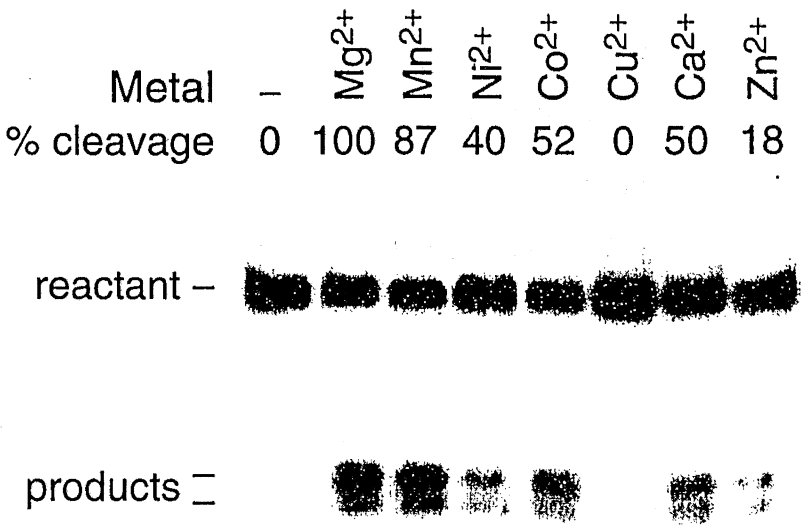
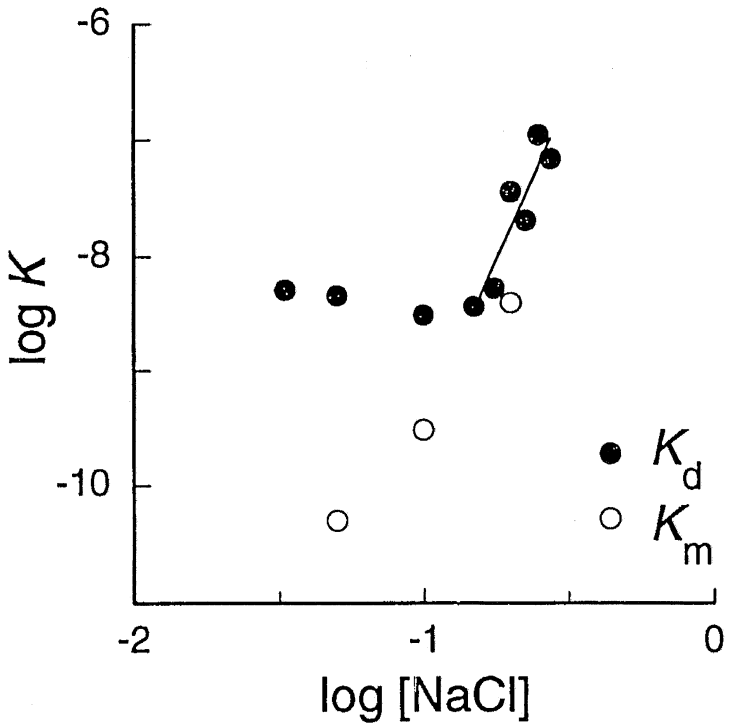

Metal
% cleavage
–
0
Mg^{2+}
100
Mn^{2+}
87
Ni^{2+}
40
Co^{2+}
52
Cu^{2+}
0
Ca^{2+}
50
Zn^{2+}
18
reactant –
products –

Figure 2.6 Binding of **I** to I-*Ppo*I in the presence of various concentrations of NaCl. The slope of the linear portion of the data suggests that approximately 6 cations are displaced from DNA upon binding of I-*Ppo*I.


-6
-8
-10
log K
-2
-1
0
log [NaCl]
K_d
K_m

Chapter 3

Degenerate Sequence Recognition by I-*Ppo*I Endonuclease

3.1 Introduction

Through the evolution of biochemical systems in organisms, nature has developed intricate specificity between enzymes and their substrates. Substrate specificity ensures that a protein catalyzes only one reaction in a biochemical pathway. Some of the most important biological functions of enzyme–substrate specificity include repairing, replicating, and maintaining the genomes of organisms. The enzymes involved in these functions recognize nucleic acids and maintain a high degree of specificity for their substrate. For example, DNA polymerase I replicates double-stranded DNA with an error rate of incorporation of only 1 per 10^6 to 10^7 basepairs (Kunkel *et al.*, 1981). The restriction enzyme, *Eco*RV, cleaves DNA at its recognition site at a rate greater than 10^6 times faster than noncognate sites; the alteration of a single base within the recognition sequence is sufficient to cripple *Eco*RV's catalytic activity (Taylor & Halford, 1989). The *lac* repressor protein, which negatively regulates transcription of genes within the lactose operon, binds its specific operator sequence 10^7 times more tightly than random nonspecific DNA (Riggs *et al.*, 1970; von Hippel *et al.*, 1975). The interactions exploited by proteins to achieve such extraordinary levels of discrimination in molecular recognition are an area of intense investigation.

Insights into the mechanisms of site-specific DNA recognition have come from the increasingly large number of structural studies of protein–DNA complexes. In the past fifteen years since the first structures of DNA-binding proteins were solved by x-ray diffraction analysis (Anderson *et al.*, 1981; McKay & Steitz, 1981; Pabo & Lewis, 1982), numerous structures of sequence-specific proteins and protein-DNA complexes have been determined. The first structures of the bacterial transcription factors, the cro and cI repressors of bacteriophage lambda, and *Escherichia coli* catabolite activator protein (CAP), suggested two important sources of sequence specificity (Steitz, 1990): first,

direct readout of the DNA basepairs, and second, indirect readout through the sugar-phosphate backbone of the DNA polymer.

Direct readout mediates sequence-specific recognition by forming complementary interactions between both side-chain and main-chain functional groups of the protein and functional groups of the DNA basepairs. These interactions include hydrogen bonding between hydrogen bond donors and acceptors on the DNA bases and protein (Figure 3.1). For example, the amino group of an adenine or cytosine base can form a hydrogen bond with a carbonyl group of a protein side-chain or main-chain. Direct readout also includes van der Waals interactions. For example, a favorable interaction can exist between the thymidine methyl group and a nonpolar amino acid side chain of the protein. Conversely, productive binding of a protein may be excluded due to steric interference by bulky groups of the basepairs in non-cognate sequences. The distinct patterns of functional groups presented by DNA basepairs allow for differentiation by DNA-binding enzymes and proteins.

Crystal structures reveal elaborate networks of direct readout interactions along the protein–DNA interface (Aggarwal *et al.*, 1988; Rosenberg, 1991). A single functional group present on one base may interact with more than one functional group on the protein, or conversely, one group present on the protein may interact with functional groups on more than one base. Additionally, some direct interactions can be mediated through water molecules. The extent to which water mediated interactions ensures specificity, however, remains controversial (Otwinowski *et al.*, 1988; Rosenberg, 1991; Kim *et al.*, 1995).

Rich and coworkers first noted that the major groove of the B-DNA helix provides a more diverse array of potential direct readout interactions than does the minor groove (Figure 3.1) (Seeman *et al.*, 1976). Predictably, many sequence-specific proteins make contact with the more variable major groove. Examples of direct readout from the crystal

structures of MAT α2 homeodomain and λ repressor with their respective recognition sequences can be seen in Figure 3.2 (Jordan & Pabo, 1988; Wolberger *et al.*, 1991). Here, the complementary hydrogen bond donors and acceptors can be seen between the protein side chains and basepair functional groups projecting into the major groove: glutamine 44 of λ repressor interacts with adenine, and arginine 54 of the homeodomain interacts with guanine. Figure 3.2 also illustrates an isoleucine residue in close contact with a cytosine base. This type of interaction may preclude productive binding at a thymidine base due to steric conflict by bulky methyl group.

Although the diversity and size (12 Å wide, 9 Å deep) of the major groove of B-DNA appear to make the major groove more amenable to making sequence-specific contacts with proteins (Blackburn & Gait, 1990), this groove is not the exclusive site of sequence-specific contacts. There are sequence-specific proteins that make significant contact in the minor groove. The transcription factor TFIID binds the sequence TATAAA(A) making contacts primarily within the minor groove (Starr & Hawley, 1991).

Additional indirect interactions such as DNA distortion aid in specific-sequence recognition. This type of recognition is an example of indirect readout. Specific DNA sequences have characteristic structures dependent on local environment. For example, the torsional flexibility of DNA allows the double helix to overwind or underwind. The homopolymer poly(dA)•poly(dT) is overwound (10 basepairs per turn as compared to the average 10.5 basepairs per turn for random DNA) (Niederweis *et al.*, 1993). This DNA overwinding narrows the minor groove with a concomitant widening of the major groove (Heinemann *et al.*, 1994). Conversely, DNA rich in (dG)•(dC) basepairs is slightly underwound (11.1 basepairs per turn) with a relatively wider minor groove (Biburger *et al.*, 1994; Heinemann *et al.*, 1994). In addition base interactions, steric clashes between functional groups, and variable stacking with adjacent bases can alter the local structure of DNA, with each sequence possessing a range of possible structures. The more readily

a sequence adopts a structure required in a specific protein–DNA complex, the less the energetic cost for binding to the specific protein. Protein–DNA complexes induce a wide range of DNA distortions such as bending (Kim *et al.*, 1989; Schultz *et al.*, 1991; Winkler *et al.*, 1993), partial unwinding (Frederick *et al.*, 1984; McClarin *et al.*, 1986), or even flipping a base completely out to the exterior of the DNA helix (Klimasauskas *et al.*, 1994). These DNA conformations and distortions can be essential elements in sequence discrimination.

The combination of direct or indirect readout used to achieve specificity by the homing endonuclease I-*Ppo*I is not currently known. Further, little is known about the modes of DNA recognition used by the any of the members of this class of endonuclease. These enzymes function *in vivo* to promote the homing of their DNA-coding sequence by creating a site-specific double-stranded break within their target DNA (Dujon, 1989). As described in Chapter 1, homing is a gene conversion event which results in the transfer of a copy of the coding sequence of the endonuclease to a cognate allele. Homing endonucleases cleave double-stranded DNA within large asymmetrical recognition sequences (15-40 basepairs) and appear to tolerate varying degrees of degeneracy within their recognition sequences (Mueller *et al.*, 1993). Much interest has been placed on the specificity of these endonucleases, as they possess potential value for genomic mapping and sequencing projects.

The I-*Ppo*I endonuclease is a 2 x 20 kDa homodimer that mediates homing of the third intron found in the gene of the large ribosomal subunit of the slime mold, *Physarum polycephalum* (Muscarella *et al.*, 1990). In the presence of Mg^{2+}, the enzyme cleaves two phosphodiester bonds within its 15-basepair homing site (Figure 3.3) (Ellison & Vogt, 1993). The cleavage sites within this sequence generate four-base 3′-hydroxyl overhangs. In the absence of Mg^{2+}, the enzyme still binds specifically to its substrate and bends its substrate 38° from linearity (Wittmayer & Raines, 1996). In addition, the

enzyme contacts regions flanking the 15-basepair recognition sequence (Ellison & Vogt, 1993). The footprint of I-*Ppo*I, using the enzymatic probe DNase and the small chemical probe dimethyl sulfate (DMS), extends over 23 basepairs. Currently, the contribution of individual bases to sequence specific recognition by I-*Ppo*I are not known.

In this study, we examine in detail the specificity of I-*Ppo*I. *In vitro,* we tested the catalytic activity of I-*Ppo*I with DNA recognition sites containing mutations that span the entire 23 basepair footprint. From this information, we found that a degree of degeneracy is tolerated within the recognition site of I-*Ppo*I. In some cases, the enzyme is capable of cleaving altered sequences. Few single point mutations debilitated the catalytic capabilities of the endonuclease. In addition, many double mutants were cleaved with equivalent or only slightly reduced efficiency from that of the wild-type sequence. Deletions or insertions within the I-*Ppo*I recognition site, however, were detrimental to catalysis indicating that proper spacing of contacts along the protein–DNA interface is critical. Finally, we observed that regions flanking the 15-basepair recognition sequence can influence the activity of I-*Ppo*I.

3.2 Materials and Methods

Bacterial Strains, Plasmids, Enzymes and Chemicals. Escherichia coli strain XL1-BLUE (*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac* [(F' *proAB*, *lac*^{q}*ZΔM15*, Tn *10* (tet^r)]) and pBluescript(SK+) were from Stratagene (La Jolla, CA). The plasmid, p42, was a gift from Promega (Madison, WI) (Muscarella *et al.*, 1990). All enzymes for the manipulation of DNA were purchased from Promega (Madison, WI). Reagents for DNA synthesis were from Applied Biosystems (Foster City, CA) except for acetonitrile, which was from Baxter Healthcare (McGaw Park, IL). Isopropyl β-D-thiogalactopyranoside (IPTG) was from Gold Biotechnology (St. Louis, MO). Ampicillin and 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-GAL) were from

Sigma Chemical (St. Louis, MO). Agarose was obtained from Life Technologies (Gaithersburg, MD). Bacto yeast extract, Bacto tryptone, and Bacto agar were obtained from Difco (Detroit, MI). All other chemicals and reagents were of commercial reagent grade or better and were used without further purification.

I-*Ppo*I, purified from a bacterial expression system, was a gift from Promega (Madison, WI). The purified protein is the larger form of the protein produced in *Physarum polycephalum* and contains 185 amino acid residues. Concentrations of the endonuclease were determined by Bradford assays using the protein assay kit from Bio Rad (Richmond, CA) using bovine serum albumin as the standard. Stock solutions of endonuclease were stored in 25 mM Tris-HCl buffer, pH 7.4, containing EDTA (10 mM), DTT (1 mM), NaCl (200 mM), and glycerol (10% v/v) and were diluted into solutions of bovine serum albumin (200 μg/ml) and glycerol (10% v/v), prior to individual assays.

General Methods. The oligonucleotides in Table 3.1 were synthesized using β-cyanoethyl phosphoramidite chemistry (Sinha *et al.*, 1984) on an Applied Biosystems Model 392 DNA-RNA synthesizer. Plasmid DNA was prepared using the Wizard Miniprep kits from Promega (Madison, WI) as instructed by manufacturer. All sequences were determined by dideoxynucleotide chain-termination sequencing using Sequenase Version 2.0 and the M13 forward primer from United States Biochemical Corp (Cleveland, OH).

Random Mutagenesis of the I-PpoI recognition site. Random mutants of the I-*Ppo*I recognition site were made by synthesizing a doped oligonucleotide of the recognition site. The oligonucleotide PW48, 5′ CGCTCTAGAGTAACTATGACTCTCTTAAGG TAGCCAAATGCCTGTCGACACCG 3′, was synthesized using a degenerate mixture of 95% of the designated base and 1.7% of each of the remaining three nucleotides at each position in the underlined region. The non-underlined sequence was synthesized with 100% of the designated base. Using the oligonucleotide PW47 as a primer, the second

strand was synthesized by treatment with Klenow. The resulting double-stranded oligonucleotide was restricted with *Sal*I and *Xba*I and ligated into the *SalI/Xba*I fragment of pBluescript(SK+). The ligation mixtures were transformed by electroporation into the *Escherichia coli* strain, XL1-BLU. The total number of recombinants obtained (8 x 10^4) was estimated by plating 10% of the liquid culture onto LB agar containing ampicillin (50 μg/ml), IPTG (0.5 mM), and X-GAL (100 μg/ml) (Ausubel *et al.*, 1989). Individual colonies were picked from this library, and liquid cultures (1.5 mL) of these colonies were grown in LB medium (Ausubel *et al.*, 1989). Plasmid DNA was prepared from these cultures, sequenced, and screened as a substrate for I-*Ppo*I.

Synthesis of symmetric recognition sites Oligonucleotides PW50–PW53 were synthesized to create two separate perfect palindromes of the I-*Ppo*I recognition site: PW50/PW51 represent a reflection of the left side of the recognition sequence and PW52/ PW53 represent a reflection of the right side of the recognition sequence. These oligonucleotides were phosphorylated by treatment with T4 polynucleotide kinase. Complementary oligonucleotides were annealed by heating to 90 °C a solution (1 μM of PW50 and PW51, or PW52 and PW53) of 10 mM Tris-HCl buffer, pH 8.0, containing NaCl (200 mM), and subsequent cooling of the solution to room temperature. The oligonucleotides were ligated into the *SalI/Xba*I site of the multiple cloning site of pBluescript(SK+), and their sequences were verified by dideoxynucleotide sequencing (CTCTCTTAAGAGAGC and CTACCTTAAGGTAGC, respectively).

Synthesis of supplementary recognition sites. Synthetic oligonucleotides PW55–PW59 were created to enhance the number of singly substituted recognition sites assayed. These oligonucleotides were used as primers in site-directed mutagenesis by the double polymerase chain reaction (Barik, 1993). Briefly, this method uses three oligonucleotide primers (a mutagenic primer, PW55–PW59; and two flanking primers, –50 Reverse and T7 promoter; Table 3.1) and two rounds of PCR performed on a template that contains

the wild-type sequence (the plasmid p42 which contains the recognition sequence of I-*Ppo*I). PCR reactions were performed using standard reaction conditions. The first round of PCR used the mutant and –50 Reverse oligonucleotides as primers to create a double-stranded product, the "megaprimer". This PCR product and the T7 promoter oligonucleotide were then used as primers in the second round of PCR again with p42 as the template. The resulting mutagenic, double-stranded oligonucleotide was purified, restricted with *Sal*I and *Xba*I, and ligated into the *SalI/Xba*I fragment of pBluescript(SK+). The sequences were verified by dideoxynucleotide sequencing.

I-PpoI endonuclease assays. Individual plasmids (approximately 250 ng) were incubated with I-*Ppo*I (0.25 ng, 2.5 ng, 25 ng) in 15 μl of 25 mM CAPS and 25 mM CHES buffer, pH 10, containing $MgCl_2$ (10 mM) and DTT (1 mM) for 1 h at 37°C. Reactions were quenched with a solution of SDS (1% w/v), EDTA (50 mM), glycerol (10% v/v), and bromophenol blue (0.01% w/v). Products were separated on a 0.75% (w/v) TAE agarose gel. Under these reaction conditions 100% of the wild-type sequence was cleaved at all enzyme:DNA ratios. The mutagenic plasmids were classified into 4 categories. Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70-95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio (Figure 3.4).

Composite analysis of entire data set. A matrix was constructed to highlight the impact of possible substitutions within the recognition site of I-*Ppo*I using the data collected for all mutant substrates assayed (Appendix). To create this summary matrix, each base substitution at each position was considered individually. Using the previously described substrate classifications, all substrates whose sequences contained a particular substitution were considered together as a subset. The base substitution was then

assigned a classification based on the substrate that was cleaved most readily. For example, if an individual base substitution was a component of three different mutant substrates of either Class 2, 3, or 4, the individual substitution was assigned the classification of 2. If an individual base substitution was a component of two different mutant substrates of either Class 3 or Class 4, the individual substitution was assigned the classification of 3. Note that this classification is a good reflection of the effect of the substitution if one of the substrates in the inspected subset was a singly substituted sequence; however, the classification is less accurate if the substrates considered only contained multiple substitutions. These classifications therefore are only considered an estimate, and they represent the best case observed for the individual mutation. Thus, this matrix only estimates the maximal effect of an individual base substitution within the I-*Ppo*I recognition sequence assuming individual mutations do not compensate for other individual mutations.

3.3 Results

We have examined the influence of DNA basepair mutations on the catalysis by I-*Ppo*I. Basepairs located within the 15 basepair recognition site of I-*Ppo*I and the outside flanking regions protected by footprinting were mutated randomly to create a population of recognition sites containing single, double, or multiple mutations. Individual members of the population were sequenced and assayed to examine the effect of the specific mutations. Of the 77 independent clones screened from this population, 10 were the wild-type sequence (13%), 11 were single mutations (14%), 20 were double mutations (26%), 19 were triple mutations (25%), 10 were quadruple mutations (13%), 2 were quintuple mutations (3%), and 5 were insertions or deletions (6%). This distribution of mutations was not as anticipated. The predicted distribution of mutations from a particular doping frequency can be determined from eq. 3.1 (Hermes *et al.*, 1990).

$$P = \frac{m!}{(m-n)!n!}\alpha^n(1-\alpha)^{(m-n)} \tag{3.1}$$

In eq. 3.1, m is the length of the doped region, α is the doping frequency, and P is the fraction of the oligonucleotides that contains n errors in the sequence. With a doping frequency of 5%, we should have observed approximately 31% wild type, 37% single mutations, 21% double mutations, 8% triple mutations. Our population was under-represented by sequences with single mutations and over-represented by those with multiple substitutions, suggesting that the actual doping frequency may have been close to 10%.

To enlarge the population of single mutations, additional clones were constructed individually using site-directed mutagenesis. These supplementary sequences are denoted with an S prefix. The final population included 102 individual clones of which 95 were independent sequences. Of this final population of 102 sequences, 10 were wild-type (10%), 23 were single mutations (23%), 30 were double mutations (29%), 19 were triple mutations (19%), 10 were quadruple mutations (10%), 2 were quintuple mutations (2%), and 8 were insertions or deletions (8%).

The plasmid substrates of the final population were screened as substrates for I-*Ppo*I and thereby grouped into categories in which catalysis by I-*Ppo*I was either unaffected (Class 1), affected slightly (Class 2), affected moderately (Class 3), or affected greatly (Class 4). A typical assay is shown in Figure 3.4. The results for all the sequences assayed are presented in the appendix. Independent clones containing the same sequence mutations displayed similar activities demonstrating the validity of the assay.

Singly substituted sites. The effect of single mutations on the activity of I-*Ppo*I is shown in Table 3.2. Nineteen independent single mutations within the I-*Ppo*I footprint region were assayed. The mutations outside the 15 basepair recognition sequence (sequences **5**, **67**, and **72**) have no effect on catalysis by I-*Ppo*I. These sequences with

flanking region mutations are cut as efficiently as the wild-type sequence. They are cut to completion at the lowest concentration of I-*Ppo*I. An additional two single mutations (sequences **60**, and **S15**) have no effect on catalysis by I-*Ppo*I (Class 1); these mutations, A(+1) to G and C(-5) to A, are within the 15 basepair recognition sequence (Table 3.2 and Figure 3.3). These two internal single mutations are also cleaved to completion even at the lowest concentration of I-*Ppo*I. Only two of the sixteen single internal mutations examined (sequences **74** and **S10**) prevent catalysis by I-*Ppo*I (Class 4). These sequences are a C(-3) to A mutation and a A(+6) to C mutation (Table 3.2 and Figure 3.3). The remaining singly substituted mutants affect cleavage by I-*Ppo*I only slightly or moderately (Class 2 or Class 3, respectively). The majority of these remaining mutants exhibit a Class 2 classification. These results indicate that degeneracy is tolerated within the 15 basepair recognition sequence of I-*Ppo*I. A summary of the results for the singly substituted mutants within the recognition site is shown in Figure 3.5.

Deletions and Insertions. In contrast to single basepair mutations, deletions and insertions within the 15 basepair recognition sequence produce poor substrates for I-*Ppo*I. Deletion and insertion mutations were inadvertently introduced into the population either during synthesis of the oligonucleotide or during cloning of the randomized population of recognition sequences (Table 3.3). Two isolated clones contained insertions (sequences **6** and **23**); six clones contained deletions (sequences **32, 40, 44, S12, S20,** and **S22**). These altered sequences were poor substrates for the enzyme (Class 4). These results indicate that the proper spacing (or registry) of the protein–DNA contacts are critical for catalysis by I-*Ppo*I.

Flanking region substitutions. The effect of flanking regions on catalysis is readily apparent from our data set (Table 3.4A and Table 3.4B). As stated above, single mutations within the flanking regions did not affect the catalytic activity of I-*Ppo*I; rather, such substrates are cleaved as well as is the wild-type sequence. Multiple mutations

within only the flanking regions, however, had varied effects on catalysis by I-*Ppo*I (Table 3.4A). Five of the seven sequences that contain multiple mutations outside the undisturbed 15 basepair recognition site display slightly (Class 2) or moderately (Class 3) reduced cleavage (sequences **54, 71,** and **15**; and sequences **4** and **39**, respectively). In addition, the effect of a single basepair change within the 15 basepair recognition sequence could be enhanced when mutated in conjunction with flanking regions (Table 3.4B). Again, five of seven single mutations examined (sequences **41, 43, 49, 14,** and **58**) have a greater reduction in cleavage activity when flanking region mutations are included within the same substrate. Interestingly, the effect of flanking regions is not absolute. Two sequences of the seven examined with additional flanking mutations (sequences **48** and **42**) do not alter the effect of the single internal mutation within the recognition site (Table 3.4B). These results reveal that mutations outside the 15 basepair recognition sequence can have a role in modulating recognition and catalysis by I-*Ppo*I.

Symmetric sites. The recognition sequence of I-*Ppo*I contains partial symmetry (Figure 3.3). Ten of the fifteen basepairs within this recognition site contribute to an imperfect palindrome. Many restriction enzymes form dimers and recognize perfect palindromes, where a monomer of the enzyme recognizes one half-site (Roberts & Halford, 1993). Because the I-*Ppo*I endonuclease also functions as a dimer (Ellison & Vogt, 1993), we examined the catalysis by I-*Ppo*I on the perfectly symmetric sites of the cognate sequence half-sites. Two symmetric oligonucleotides were constructed: one reflecting the left-half of the site about the dyad of symmetry, and the other reflecting the right half about the dyad of symmetry (Figure 3.3). Both of these substrates containing two mutations are cleaved as well as is the wild-type sequence (Class 1) (data not shown). This result is consistent with a dimeric I-*Ppo*I endonuclease recognizing the partial symmetry of the recognition site.

Multiply substituted sites. Three types of effects are observed with the multiply substituted substrates: tolerance, cumulative effects, and non-cumulative effects (Table 3.5A, Table 3.5B, and Table 3.5C, respectively). First, the tolerance of I-*Ppo*I for degeneracy within its recognition sequences is readily apparent with multiply substituted substrates. Table 3.5A shows that numerous sequences with multiple mutations are cleaved as well as is the wild-type sequence (Class 1) or only slightly slower than is the wild-type sequence (Class 2). This data subset suggests that the enzyme–substrate complex is capable of adjusting to compensate for any interactions that may be lost because of mutations within the substrate. The protein appears to be able maintain or maximize favorable interactions and reduce or minimize unfavorable interactions.

Second, the effect of multiple mutations within the same substrate can be cumulative. Table 3.5B shows sequences with multiple mutations in which the effect on the activity of I-*Ppo*I from multiple mutations is larger than that from any of the single mutations. In this data subset, I-*Ppo*I exhibits a graded discrimination between wild-type and mutant sequences. Single mutations alone can have no effect or only a slight effect on cleavage, but in combination with an additional mutation, cleavage is slowed to a greater extent. For example, the double mutant that contains an A(+1) to T mutation and a C(-3) to T mutation (sequence **S25**), is a Class 3 mutant, but the single C(-3) to T mutation (sequence **S23**) is a Class 2 mutant, and two sequences with an A(+1) to T mutation (sequences **46** and **62**) are Class 1 mutants. The sequences in Table 3.5B are examples of mutations that are cumulative; together the mutations exert a greater effect on catalysis by I-*Ppo*I than they do alone.

The final set of multiple mutations are shown in Table 3.5C. The effect of these mutations are non-cumulative; the mutations combined within the same substrate exert an effect no greater than the most debilitating mutation. For example, the triple mutant which contains a T(-6) to C mutation, an A(+1) to G mutation, and an A(+10) to C

mutation (sequence **76**), is a Class 3 mutant. This sequence is cleaved no more slowly than the single T(-6) to C mutation (sequence **S19**). Despite the presence of mutations additional to the T (-6) to C, the cleavage of the substrate is not slowed to a greater extent.

Together the data in Table 3.2 through Tables 3.5, show that sequence recognition by I-*Ppo*I is complex. Many single and multiple mutations are tolerated within the recognition sequence. When the entire set of sequences are examined, the majority of the positions within the recognition sequence show tolerance for base substitutions (Figure 3.6). The presence of particular mutations in a substrate do not prevent catalysis by I-*Ppo*I. However, the effect of multiple mutations may or may not be cumulative. Therefore, predicting *a priori* the effect of multiple mutations on the catalysis by I-*Ppo*I is challenging.

3.4 Discussion

This study is the first extensive mutational analysis of the recognition site of I-*Ppo*I. We have found that, *in vitro*, I-*Ppo*I has a relaxed specificity greater than previously thought (Lowery *et al.*, 1992). Sequences with single, double, and multiple basepair substitutions within the homing site of I-*Ppo*I are acceptable substrates for the enzyme (Table 3.2 and Table 3.5A). The tolerance for degeneracy by I-*Ppo*I within its recognition sequence also appears in other homing endonucleases. Studies of the homing endonucleases, I-*Sce*I, I-*Sce*II, I-*Ceu*I, I-*Cre*I, and I-*Tev*I demonstrate that these enzymes tolerate substitutions within their large asymmetric recognition sequence (Colleaux *et al.*, 1988; Marshall & Lemieux, 1992; Wernette *et al.*, 1992; Byrk *et al.*, 1993; Dürrenberger & Rochaix, 1993). Limited tolerance for degeneracy by these endonucleases is not surprising in light of their function in *vivo*. They propagate their coding sequences to

new genetic locations. If these enzymes were to specific they would never would never migrate; however, if they were to promiscuous they would kill its host.

The level of degeneracy accepted by the different homing endonucleases span the spectrum from high to low specificity. The enzymes I-*Sce*I and I-*Ceu*I represent the most specific homing endonucleases. I-*Sce*I cleavage activity is affected severely by 14 single basepair substitutions and affected moderately by an additional 11 single basepair substitutions within its 18 basepair recognition sequence (Colleaux *et al.*, 1988). I-*Ceu*I cleavage activity is eliminated by 11 single mutations within its 19 basepair recognition sequence (Marshall & Lemieux, 1992). The homing endonucleases, I-*Sce*II and I-*Cre*I, are less stringent. Of the 36 single basepair substitutions examined within the 18 basepair recognition sequence of I-*Sce*II, most were tolerated by the enzyme. For example, only 9 mutations blocked cleavage moderately and only 3 blocked cleavage completely (Wernette *et al.*, 1992). Limited analysis of the I-*Cre*I endonuclease reveals that this enzyme cleaves its recognition site with a level of specificity similar to that of I-*Sce*II (Dürrenberger & Rochaix, 1993). Single and multiple mutations are tolerated by this homing endonuclease. The more permissive endonucleases are the T4 phage endonucleases, I-*Tev*I and I-*Tev*II. No single basepair within the 48 basepairs spanning the homing site of I-*Tev*I is essential for cleavage (Byrk *et al.*, 1993). Our data suggests I-*Ppo*I has intermediate specificity, falling between I-*Sce*I and I-*Tev*I. I-*Ppo*I appears to be slightly more specific than I-*Sce*II.

Two single substitutions of the 17 examined mutations within the 15 basepair recognition sequence of I-*Ppo*I, C(-3) to A and A(+6) to C, were shown to debilitate the cleavage activity by I-*Ppo*I (Class 4; Table 3.2). These two positions (+6 and -3) are presumably key regions for I-*Ppo*I recognition. Position +6 is partially protected in the minor groove during footprinting experiments with the small chemical probe, DMS, and the neighboring G(+7) is strongly protected in the major groove (Ellison & Vogt, 1993).

This information in conjunction with the cleavage data on the sequence with a single mutation at position +6 (sequence **74**), suggests this basepair is an important protein–DNA contact. Interestingly, I-*Ppo*I does not appear to contact closely the -3 position, as the complementary base for position C(-3), G(-3), is not protected during footprinting experiments with DMS (Ellison & Vogt, 1993). In fact, the G(-3) position when complexed with I-*Ppo*I displays a partial hypersensitivity to DMS. The hypersensitivity of G(-3) suggests that there is a distortion of the DNA helix in this position. Perhaps the substitution of the CG(-3) basepair with an AT basepair prevents the proper formation of a requisite structure alteration by I-*Ppo*I. This result suggests that I-*Ppo*I relies to some extent on indirect readout for its specificity.

The flanking regions outside the 15 basepair recognition sequence (Figure 3.3) can modulate cleavage by I-*Ppo*I. Some sequences with multiple mutations in only the flanking regions were slower substrates for I-*Ppo*I (Table 3.4A). Likewise, the cleavage rate of some sequences with single internal mutations were slowed to a greater extent when combined with flanking mutations (Table 3.4B). Flanking sequences are known to have similar modulating effects in other DNA-binding proteins. For example, cleavage by the site-specific endonuclease, HO endonuclease, is shaped by regions outside the core sequence (Nickoloff *et al.*, 1990). Most point mutations outside this core region of approximately eight basepairs have little effect on activity, whereas multiple changes outside this core region severely diminish cleavage by the HO endonuclease. Similarly, the zinc-finger DNA-binding protein, YY1, displays considerable heterogeneity in the regions flanking its conserved core sequence (Hyde-DeRuyscher *et al.*, 1995). Furthermore, contacts located outside the *Eco*RI recognition site can influence the rate of cleavage for the proximal half-site by the *Eco*RI restriction endonuclease (Van Cleve & Gumport, 1992). The footprint of the I-*Ppo*I endonuclease shows that flanking regions are protected in the I-*Ppo*I–substrate complex (Ellison & Vogt, 1993). From our data, the

interactions at these positions appear to influence the cleavage efficiency by the I-*Ppo*I endonuclease.

Few single mutations were observed to debilitate catalysis by I-*Ppo*I. A majority of the individual single mutations within the 15 basepair recognition sequence only had a slight effect on the cleavage by I-*Ppo*I (Table 3.2). These results suggest that individual basepair interactions make only small contributions to specificity. Maintaining interactions with each DNA basepair within the recognition sequence apparently is not necessary for catalysis by I-*Ppo*I. Remaining protein–DNA interactions can compensate for any lost contacts between the enzyme and substrate. Alternatively, these results suggest a more flexible, complementary I-*Ppo*I–substrate interface in which the complex adapts to form substitute interactions that contribute favorably to catalysis. Analogous adaptations have been proposed to occur within the protein–DNA complexes of *lac* repressor and mutant operator sequences (Mossing & Record, 1985), and of *Eco*RI endonuclease and analog-containing substrates (Jen-Jacobson *et al.*, 1991).

The effect of multiple mutations on catalysis by I-*Ppo*I is varied. Multiple mutations can have an cumulative or non-cumulative effect. For example, single mutations that alone have little or no effect can exhibit a greater effect on catalysis by I-*Ppo*I when combined on the same substrate (Table 3.5B). This cumulative effect of multiple substitutions would suggest a more global mode of sequence recognition. Still, other single mutation combinations can result in no alteration of the effect on cleavage (Table 3.5C). These combinations of mutations have no greater effect than that of the most debilitating single mutation alone.

Together these results exemplify a clear difference between I-*Ppo*I endonuclease and the type II restriction endonucleases. I-*Ppo*I does not maintain the strict specificities of restriction enzymes. Instead, I-*Ppo*I appears to recognize its substrates in a manner more reminiscent of transcription factors, which maintain a sequence preference rather than a

strict sequence requirement (Mossing & Record, 1985). I-*Ppo*I and other homing endonucleases, therefore, represent a distinct class of DNA-binding proteins with the complex substrate specificities of transcription factors superimposed on the catalytic capabilities of restriction endonucleases.

Table 3.1: Synthetic Oligonucleotides Used for Creation of Mutant Substrates for I-*Ppo*I

Oligo	Sequence
PW47	CGCTCTAGAGTAACT
PW48[a]	CGCTCTAGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTGT CGACACCG
PW50	CTAGAGTAACTATGACTCTCTTAAGAGAGCCAAATGCCTG
PW51	TCGACAGGCATTTGGCTCTCTTAAGAGAGTCATAGTTACT
PW52	CTAGAGTAACTATGACTACCTTAAGGTAGCCAAATGCCTG
PW53	TCGACAGGCATTTGGCTACCTTAAGGTAGTCATAGTTACT
PW55[b]	GAHTCATAGTTACTAATTA
PW56[b]	AGBGTCATAGTTACTAATT
PW57[b]	GAHAGTCATAGTTACTAAT
PW58[b]	AGBGAGTCATAGTTACTAAT
PW59[b]	AAHAGAGTCATAGTTACTAAT
–50 Reverse	TTGTGAGCGGATAACAATTTC
T7 promoter	TAATACGACTCACTATAG

[a] The underlined region was synthesized with the degenerate mixture described in text.
[b] H represents an equal mixture of A, C, and T. B represents an equal mixture of G, T, and C.

Table 3.2: Effect of Single Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
5	ATGA CTCTCTTAAGGTAGC C**G**AA	1
67	ATG**T** CTCTCTTAAGGTAGC CAAA	1
72	ATGA CTCTCTTAAGGTAGC CA**G**A	1
12	ATGA CTCTCTTAAGGTAG**G** CAAA	2
8	ATGA CTCTCTTAAGGTA**T**C CAAA	2
74	ATGA CTCTCTTAAGGT**C**GC CAAA	4
57	ATGA CTCTCTTAAG**C**TAGC CAAA	2
36	ATGA CTCTCTTAA**T**GTAGC CAAA	2
60	ATGA CTCTCTT**G**AGGTAGC CAAA	1
30	ATGA CTCTC**C**TAAGGTAGC CAAA	3
S10	ATGA CTCT**A**TTAAGGTAGC CAAA	4
S23	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S24	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S17	ATGA CTC**G**CTTAAGGTAGC CAAA	1.5[c]
S14	ATGA CT**T**TCTTAAGGTAGC CAAA	2
S15	ATGA CT**A**TCTTAAGGTAGC CAAA	1
S11	ATGA C**A**CTCTTAAGGTAGC CAAA	2.5[c]
S18	ATGA C**A**CTCTTAAGGTAGC CAAA	2.5[c]
S19	ATGA C**C**CTCTTAAGGTAGC CAAA	3
S4	ATGA **A**TCTCTTAAGGTAGC CAAA	2
S5	ATGA **T**TCTCTTAAGGTAGC CAAA	2.5[c]
S6	ATGA **A**TCTCTTAAGGTAGC CAAA	2
75	ATGA **A**TCTCTTAAGGTAGC CAAA	2

[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

[c] When the results of two individual assays varied the average of the two assays is given.

Table 3.3: Effect of Basepair Insertions and Deletions in the Substrate on the Catalysis by I-*Ppo*I

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
6	ATGA CTCTCTTAA(**T**)GGTAGC C**T**AAA	4
23	ATGA CTCTC(**T**)TTAAGGTAGC CAAA	4
32	ATGA CTCTCTTAΔGGTAGC C**G**AA	4
40	ATGA CΔCTCTTAAGGTAGC CAAA	4
44	ATGA CTC**A**CΔTAAGGTAGC CAAA	4
S12	ATGA CTΔTCTTAAGGTAGC CAAA	4
S20	ATGA CTC**A**CTΔAAGGTAGC CAAA	4
S22	ATGA CTC**A**CΔTAAGGTAGC CAAA	4

[a] Inserted bases are highlighted with bold print in parentheses. Deleted bases are represented by a Δ. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

Table 3.4A: Effect of Mutations in the Sequence Flanking the Substrate on the Catalysis by I-*Ppo*I

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
34	**T**TGA CTCTCTTAAGGTAGC C**T**AA	1
55	ATG**T** CTCTCTTAAGGTAGC CAA**C**	1
15	ATG**T** CTCTCTTAAGGTAGC CAGA	2
54	AT**C**A CTCTCTTAAGGTAGC **T**AAA	2
71	A**G**GA CTCTCTTAAGGTAGC **T**AA**T**	2
4	**G**TGA CTCTCTTAAGGTAGC **A**AAG	3
39	**T**TG**C** CTCTCTTAAGGTAGC CAGA	3
Single basepair mutations		
5	ATGA CTCTCTTAAGGTAGC C**G**AA	1
67	ATG**T** CTCTCTTAAGGTAGC CAAA	1
72	ATGA CTCTCTTAAGGTAGC CA**G**A	1

[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

Table 3.4B: Effect of Single Internal Mutations with Flanking Sequence Mutations on the Catalysis by I-*Ppo*I

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
Effect by flanking sequence mutations		
41	ATGA CTCTCTT**G**AGGTAGC CAA**C**	2
43	ATG**C** CTCTCTT**G**AGGTAGC CAA**C**	4
60	ATGA CTCTCTT**G**AGGTAGC CAAA	1
49	ATG**T** CTC**G**CTTAAGGTAGC CAAA	3
S17	ATGA CTC**G**CTTAAGGTAGC CAAA	1.5[c]
14	A**G**GA **T**TCTCTTAAGGTAGC CAAA	4
S5	ATGA **T**TCTCTTAAGGTAGC CAAA	2.5[c]
58	**G**TGA CTCTCTTAAG**C**TAGC **AAG**A	4
57	ATGA CTCTCTTAAG**C**TAGC CAAA	2
No effect by flanking sequence mutations		
48	A**A**GA CT**A**TCTTAAGGTAGC CAA**C**	1
S15	ATGA CT**A**TCTTAAGGTAGC CAAA	1
42	ATG**G** CTCTCTTAAGGTAG**G** CA**T**A	2
12	ATGA CTCTCTTAAGGTAG**G** CAAA	2

[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

[c] When the results of two individual assays varied the average of the two assays is given.

Table 3.5A: Effect of Multiple Basepair Mutations in the Substrate on the Catalysis by I-*Ppo*I: Substrates that Exhibit No or Minimal Effect

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
46	ATGA CGCTCTTTAGGTAGC CAAA	1
48	AAGA CTATCTTAAGGTAGC CAAC	1
42	ATGG CTCTCTTAAGGTAGG CATA	2
2	GTGT CTCTCTTCAGGTAGC CTTA	1
35	AGGA CTTTCTAAAGGTAGC CAAA	2
9	ATGA CTCTCTTAAGGCATC CATA	2
31	ATGA CTCGCTTAAGGCAGT CAAA	2
45	ATGA CTATTTAAAGGTAGC CAAG	2
33	ATGA CTCTCGTAAGGTAGC CATA	1
62	ATGA CTCTCTTTAGGTAGC CACA	1
41	ATGA CTCTCTTGAGGTAGC CAAC	2
66	ATGA CTCACTTAAGGTAGC CAAG	2

[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

Table 3.5B: Effect of Multiple Basepair Mutations in the Substrate on the Catalysis by I-*Ppo*I: Substrates that Exhibit a Cumulative Effect

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
S25	ATGA CTCT**T**TT**T**AGGTAGC CAAA	3
46	ATGA C**G**CTCTT**T**AGGTAGC CAAA	1
62	ATGA CTCTCTT**T**AGGTAGC CA**C**A	1
S23	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S2	ATGA **T**TC**A**CTTAAGGTAGC CAAA	4
S3	ATGA **A**TC**A**CTTAAGGTAGC CAAA	4
S5	ATGA **T**TCTCTTAAGGTAGC CAAA	2.5[c]
S6	ATGA **A**TCTCTTAAGGTAGC CAAA	2
66	ATGA CTC**A**CTTAAGGTAGC CAA**G**	2
28	ATGA CTC**A**CTTAAGGTAGC CAA**C**	3
73	ATGA CTCTCT**A**AAGGTA**T**C CAAA	4
8	ATGA CTCTCTTAAGGTA**T**C CAAA	2
35	A**G**GA CT**T**TCT**A**AAGGTAGC CAAA	2
45	ATGA CT**A**T**T**T**A**AAGGTAGC CAA**G**	2
S21	ATGA CTC**G**CT**A**AAGGTAGC CAAA	3
S17	ATGA CTC**G**CTTAAGGTAGC CAAA	1.5[c]
45	ATGA CT**A**T**T**T**A**AAGGTAGC CAA**G**	2
35	A**G**GA CT**T**TCT**A**AAGGTAGC CAAA	2

[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

[c] When the results of two individual assays varied the average of the two assays is given.

Table 3.5C: Effect of Multiple Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I: Substrates that Exhibit a Non-Cumulative Effect

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
76	ATGA C**C**CTCTT**G**AGGTAGC C**C**AA	3
60	ATGA CTCTCTT**G**AGGTAGC CAAA	1
S19	ATGA C**C**CTCTTAAGGTAGC CAAA	3
9	ATGA CTCTCTTAAGG**C**A**T**C CA**T**A	2
8	ATGA CTCTCTTAAGGTA**T**C CAAA	2

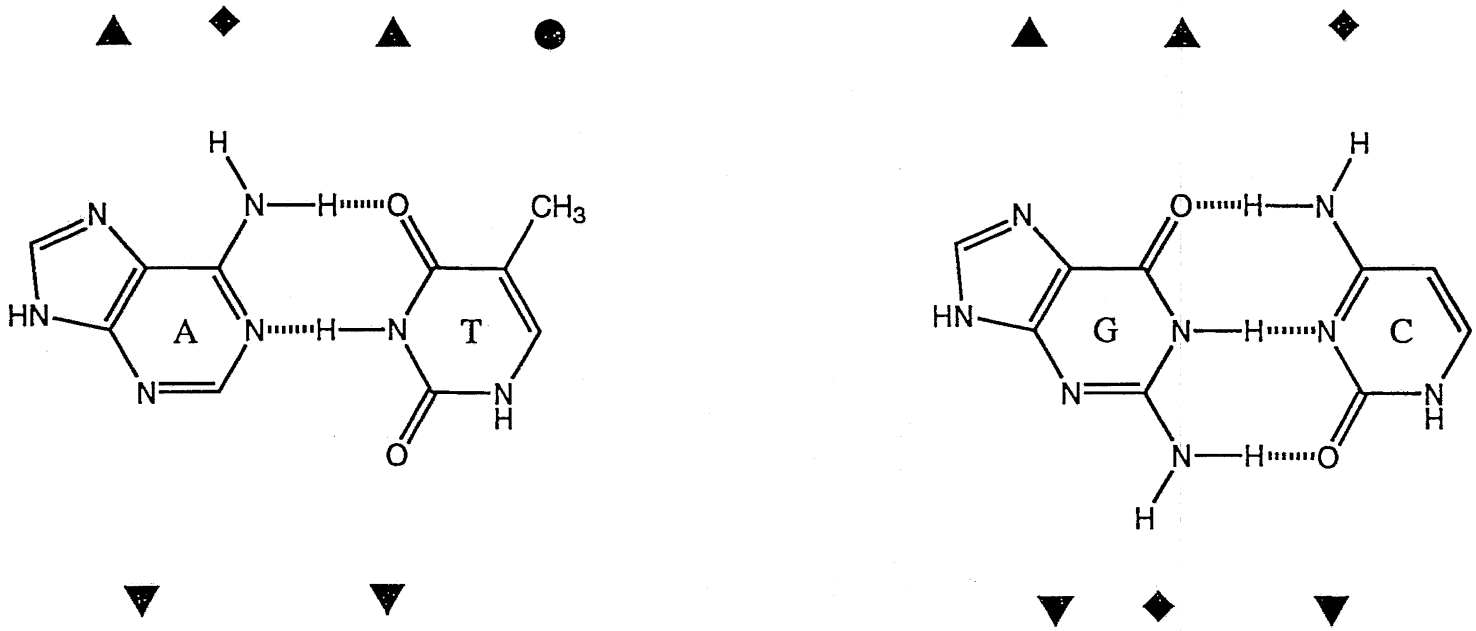
[a] Substituted bases are highlighted with bold print. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

[c] When the results of two individual assays varied the average of the two assays is given.

Figure 3.1 Functional groups of Watson-Crick basepairs presented in the major and minor grooves of B-DNA. A more diverse array of functional groups are available in the major groove versus the minor groove for sequence discrimination (Seeman *et al.*, 1976). The symbols, ▲ and ◆, represent hydrogen bond acceptors and donors, respectively. The symbol, ●, represents the bulky methyl group of thymidine.

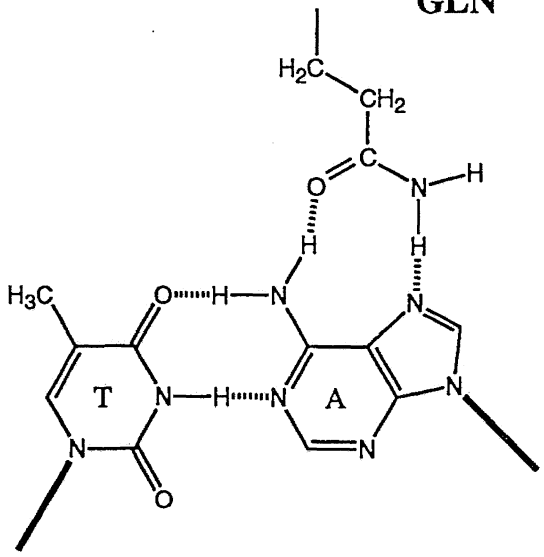
Major Groove



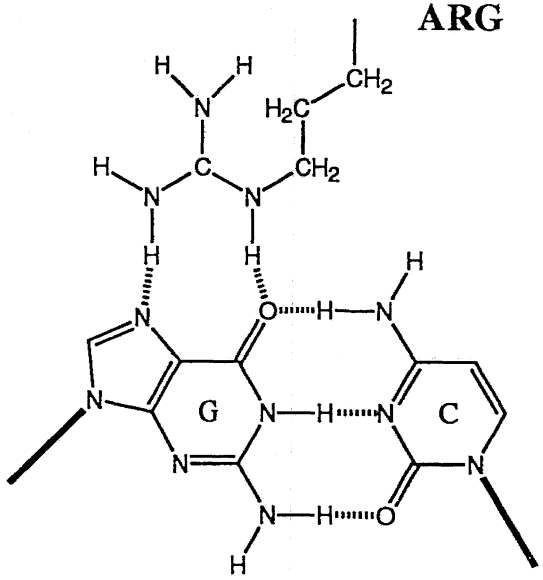
Minor Groove

Figure 3.2 Direct readout of DNA basepairs. Examples of direct readout from the structures of crystalline MAT α2 homeodomain and λ repressor with their respective recognition sequences (Jordan & Pabo, 1988; Wolberger *et al.*, 1991). Glutamine 44 of λ repressor interacts with adenine, and arginine 54 of the homeodomain interacts with guanine. Also, an isoleucine residue in close contact with a cytosine base showing how such an interaction may preclude productive binding at a thymidine base due to steric conflict by bulky methyl group.

GLN


H2C
CH2
C
O
N
H
H
H3C
O
H
N
N
T
N
H
N
A
N
N
N
O

ARG


H
H
N
H2C
CH2
H
N
C
N
CH2
H
H
N
O
H
N
H
N
G
N
H
N
C
N
N
N
H
O
H

ILE

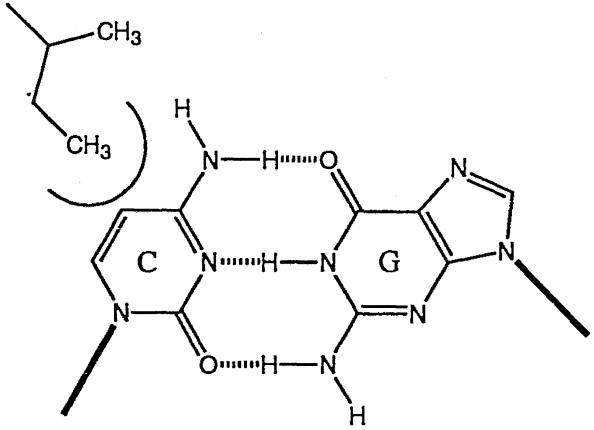

CH3
CH3
H
N
H
O
N
C
N
H
N
G
N
N
N
O
H
N
H

Figure 3.3 Homing site of I-*Ppo*I including flanking regions. The DNA bases are numbered about the dyad of pseudosymmetry. The bases that contribute to the symmetry are in bold. The central 15 basepairs comprise the recognition site as determined by deletion analysis (Ellison & Vogt, 1993). The bases outside the central 15 basepairs are the flanking regions protected in the I-*Ppo*I footprint. The arrow denotes the location of the Pp LSU3 intron insertion, and the site cleaved is indicated by a line. All 23 basepairs represent potential sites of substitution in our studies.

		-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8		
5'	ATGA	**C**	**T**	C	T	**C**	**T**	**T**	**A**	**A**	**G**	G	T	**A**	**G**	C	CAAA	3'
3'	TACT	**G**	**A**	G	A	**G**	**A**	**A**	**T**	**T**	**C**	C	A	**T**	**C**	G	GTTT	5'

Intron Insertion Site

Figure 3.4 Activity assay of mutant substrates with the I-*Ppo*I endonuclease. Plasmid substrates were incubated with increasing concentrations of enzyme (1 nM, 10 nM, 100 nM) in a 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM) and DTT (1 mM) for 1 h at 37°C. Lane 1 is λ DNA digested with *Hind*III, lanes 2-5 is the wild-type substrate; lanes 6-9 is sequence **62**, a class 1 substrate; lanes 10-13 is sequence **64**, a class 3 substrate; and lanes 14-17 is sequence **65**, a class 4 substrate. Lanes 2, 6, 10, and 14 have no endonuclease; lanes 3, 7, 11, and 15 have 1 nM endonuclease; lanes 4, 8, 12, and 16 have 10 nM endonuclease; and lanes 5, 9, 13, and 17 have 100 nM endonuclease. The different forms of the plasmid are indicated on the right: n. indicates nicked DNA, l. indicates linear DNA, and s.c. indicates supercoiled DNA.

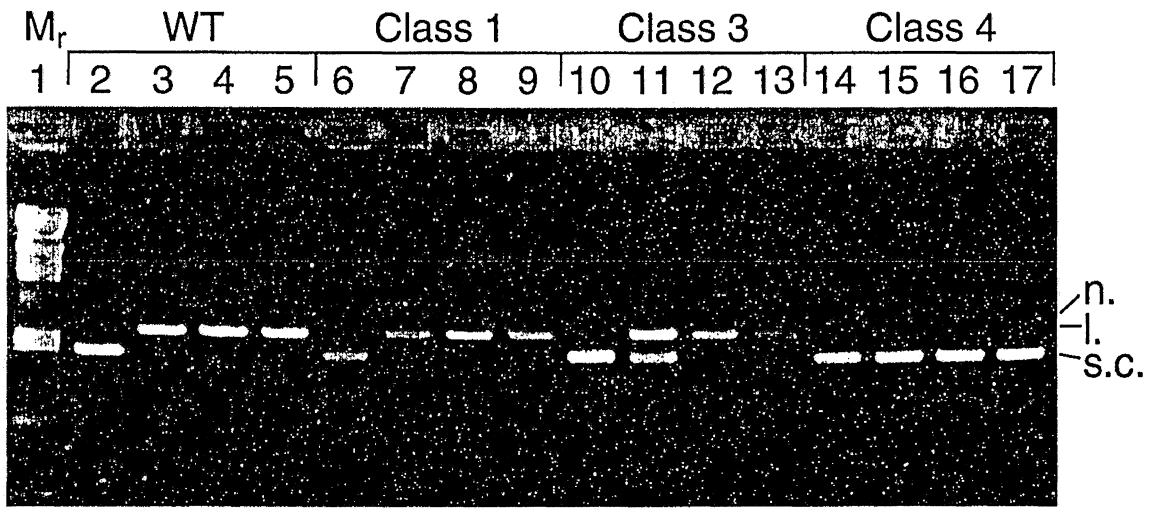

Mr
WT
Class 1
Class 3
Class 4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
n.
l.
s.c.

Figure 3.5 Summary of sequences examined with single base substitutions. The data in Table 3.2 is summarized as a matrix. The wild-type recognition sequence is given (top). The three possible base substitutions are shown (left). Black boxes represent the wild-type sequence. White boxes indicate base substitutions for which no singly substituted sequence was found. Colored regions represent the results for the base substitutions for which singly substituted sequences were found. Analogous to a traffic light, red boxes represent substitutions that "stopped" catalysis by I-*Ppo*I; green boxes represent substitutions that allowed catalysis by I-*Ppo*I to "go". Class 1 substrates are dark green; Class 2 substrates are light green; Class 3 substrates are light red; and, Class 4 substrates are dark red.

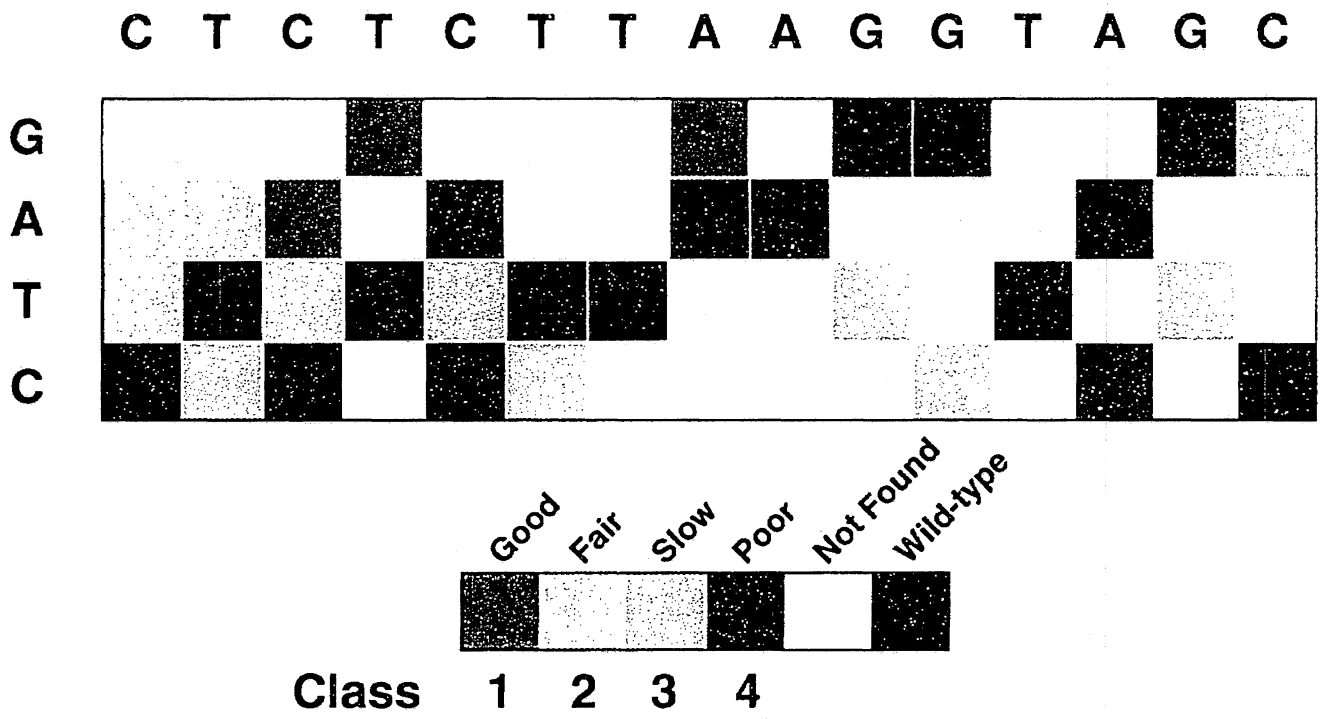
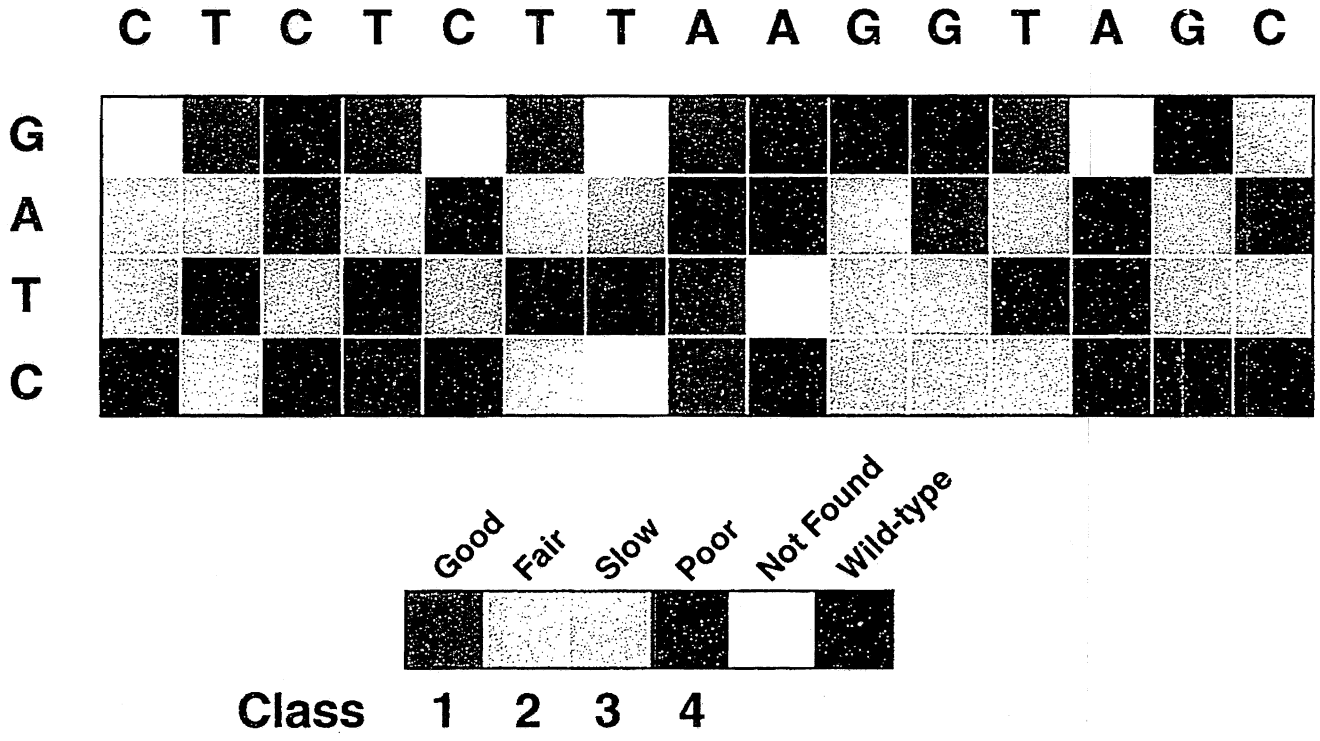

C T C T C T T A A G G T A G C
G
A
T
C
Good
Fair
Slow
Poor
Not Found
Wild-type
Class 1 2 3 4

Figure 3.6 Summary of all sequences examined. The data in the appendix is summarized as a matrix. The wild-type recognition sequence is given (top). The three possible base substitutions are shown (left). Black boxes represent the wild-type sequence. White boxes indicate base substitutions for which no sequence containing that mutation was found. Colored regions represent the results for base substitutions for which sequences containing that mutation were found. Analogous to a traffic light, red boxes represent substitutions that "stopped" catalysis by I-*Ppo*I; green boxes represent substitutions that allowed catalysis by I-*Ppo*I to "go". Class 1 substrates are dark green; Class 2 substrates are light green; Class 3 substrates are light red; and, Class 4 substrates are dark red. Note, the classification of individual bases from data with multiple substitutions are estimated to be only a maximal effect. For bases represented in green (good or fair substrates), this estimate is likely to be reasonable. For bases represented in red (slow or poor substrates), this estimate is less likely to be reasonable unless the substitution was analyzed as a singly substituted sequence (Figure 3.5). If the base represented in red was only analyzed in a substrate with multiple mutations, the "stopped" catalysis by I-*Ppo*I could be due to other mutations within the sequence or a cumulative effect (Table 3.5B).


C T C T C T T A A G G T A G C
G
A
T
C
Good
Fair
Slow
Poor
Not Found
Wild-type
Class 1 2 3 4

Chapter 4

Production, Purification, and Characterization of I-*Ppo*I Mutants: Searching for Active-Site Residues

4.1 Introduction

The homing endonuclease, I-*Ppo*I, initiates intron homing by cleaving the extrachromosomal nuclear DNA of the slime mold, *Physarum polycephalum* within I-*Ppo*I's large recognition site (Muscarella *et al.*, 1990):

```
C T C T C T T A A↓G G T A G C
G A G A G↑A A T TC C A T C G
```

Intron homing is a gene conversion process that results in the duplication of the intron encoding I-*Ppo*I into other cognate alleles (Dujon, 1989). The I-*Ppo*I endonuclease is a member of a class of enzymes that share common biochemical properties (Mueller *et al.*, 1993). Most of the eukaryotic homing endonucleases cleave within large, asymmetric recognition sequences to generate 4-basepair 3′-OH overhangs, and require Mg^{2+} ion as a cofactor. I-*Ppo*I, however, does not share a common structural motif with these enzymes, that is the amino acid sequence LAGLI-DADG (Chapter 1). Mutagenesis studies of the homing endonucleases PI-*Sce*I and PI-*Tli*I have identified the conserved acidic residue within this amino acid motif (LAGLI-DA<u>D</u>G) as being essential for catalysis by these enzymes (Hodges *et al.*, 1992; Gimble & Stephens, 1995).

The study of amino acid motifs has been successful in identifying active-site amino acids in other endonucleases such as the type II restriction endonucleases. These enzymes also require Mg^{2+} ion as a cofactor for specific cleavage of DNA. The active site residues of several restriction endonucleases have been located through sequence and structural comparisons. The amino acid motif P-(D/E)-X_n-(D/E)-X-K describes the active site of both *Eco*RI and *Eco*RV (Table 4.1) (Thielking *et al.*, 1991; Selent *et al.*, 1992; Grabowski *et al.*, 1995). X-ray crystallographic analysis of these enzymes, reveal that the four amino acids Pro90, Asp91, Glu111, and Lys113 of *Eco*RI, and the four

amino acids Pro73, Asp74, Asp90, and Lys92 of *Eco*RV, are located near the scissile phosphodiester bond in their respective enzyme–substrate complexes (Figure 4.1) (Kim *et al.*, 1990; Winkler *et al.*, 1993; Kostrewa & Winkler, 1995). Analysis of the structures of *Pvu*II and *Bam*HI with their respective substrate reveals that the amino acid residues Asp58, Glu68, and Lys70 of *Pvu*II, and the amino acid residues Asp94, Glu111, and Glu113 of *Bam*HI are similarly positioned about the scissile phosphodiester bond (Cheng *et al.*, 1994; Newman *et al.*, 1995). In these two enzyme sequences however, the proline residue is missing, and in the case of *Bam*HI, the basic residue is replaced by an acidic residue Glu113 (Table 4.1).

Amino acid residues critical for catalysis by the restriction endonucleases *Fok* I and *Nae*I were identified without the structures of these enzymes (Waugh & Sauer, 1993; Holtz & Topal, 1994). The amino acid residues Asp450 and Asp467 of *Fok*I were identified as components of the P-(D/E)-X_n-(D/E)-X-K motif (Table 4.1). Mutagenesis of these residues to alanine identified Asp450 and Asp467 as residues critical for catalysis by *Fok*I but not essential for specific substrate recognition (Waugh & Sauer, 1993). Alternatively, random mutagenesis of the gene encoding *Nae*I has identified Glu70 and Asp95 as essential residues for catalysis (Holtz & Topal, 1994). These essential amino acids are components of the same amino acid motif within the *Nae*I sequence (Table 4.1). Although the amino acid sequences of the discussed restriction enzymes exhibit no significant sequence homology, there is a striking similarity in the amino acid signature of their active sites.

The homing endonuclease, I-*Ppo*I shows no significant sequence homology with other homing endonucleases or restriction endonucleases. Nonetheless, the sequence of I-*Ppo*I does contain amino acids which apparently correspond to the active-site motif P-(D/E)-X_n-(D/E)-X-K (Chapter 2 and Table 4.1). In order to test the functional importance of the amino acid residues of I-*Ppo*I that coincide with this motif, Asp109,

Asp140 and Lys142 were changed to alanine by site-directed mutagenesis, and the effects of these replacements on the activity of I-*Ppo*I were analyzed. The results of these experiments suggest that amino acid Lys142 is part of the active site of I-*Ppo*I; however, amino acid residues Asp109 and Asp140 do not appear to be essential catalytic residues. All three of these mutants bind the I-*Ppo*I recognition sequence, yet only the K142A mutant is catalyticly deficient. The two aspartate to alanine mutants retain at least 10% of the wild-type activity. Despite the similarities of the reactions catalyzed by both homing endonucleases and restriction endonucleases, the active-site of I-*Ppo*I may contain a distinct architecture to catalyze the site-specific cleavage of DNA.

4.2 Materials and Methods

Materials. Escherichia coli strain BL21 (DE3) (F^- ompT r_B-m_B-; (Studier & Moffatt, 1986)) was acquired from Novagen (Madison, WI). *Escherichia coli* strain CJ236 and helper phage M13K07 were from Bio-Rad (Richmond, CA). All enzymes for the manipulation of DNA were from Promega (Madison, WI). Reagents for DNA synthesis were from Applied Biosystems (Foster City, CA) except for acetonitrile, which was from Baxter Healthcare (McGaw Park, IL). Isopropyl β-D-thiogalactopyranoside (IPTG), was from Gold Biotechnology (St. Louis, MO). Ampicillin, sodium deoxycholate, guanidine-HCl, and poly(dIdC)·poly(dIdC) were from Sigma Chemical (St. Louis, MO). Agarose was from Life Technologies (Gaithersburg, MD). Bacto yeast extract, Bacto tryptone, and Bacto agar were obtained from Difco (Detroit, MI). All other chemicals and reagents were of commercial reagent grade or better and were used without further purification. Wild-type I-*Ppo*I, purified from a bacterial expression system, and the plasmid, p42 (Muscarella *et al.*, 1990), was a gift from Promega (Madison, WI). The purified protein is the larger form of the protein produced in *Physarum polycephalum* and contains 185 amino acid residues.

General Methods. Protein concentrations were determined using an extinction coefficient at 280 nm of 36,492 $M^{-1}cm^{-1}$ as determined for the wild-type protein with the method of Gill and von Hippel (Gill & von Hippel, 1989). This measure of protein concentration is consistent with results obtained from Bradford assays using the protein assay kit from Bio Rad (Richmond, CA) with bovine serum albumin as the standard. Stock solutions of the endonucleases were stored in 25 mM Tris-HCl buffer, pH 7.4, containing EDTA (10 mM), DTT (1 mM), NaCl (200 mM), and glycerol (10% v/v) and were diluted into solutions containing bovine serum albumin (200 μg/ml) and glycerol (10% v/v) prior to individual assays. Proteins were separated by PAGE performed in the presence of SDS (0.1% w/v), as described (Ausubel *et al.*, 1989). Gels were fixed and stained by washing with aqueous methanol (40% v/v) containing acetic acid (10% v/v) and Coomassie brilliant blue (0.1% w/v). Ultraviolet and visible absorbance measurements were made with a Cary Model 3 spectrophotometer (Varian, Sugar Land, TX). Circular dichroism spectra for wild-type, D109A, D140A, and K142A I-*Ppo*I were measured using a Cary 60 circular dichroism spectrometer from Olis (Bogart, GA). The oligonucleotides were synthesized using β-cyanoethyl phosphoramidite chemistry (Sinha *et al.*, 1984) on an Applied Biosystems Model 392 DNA-RNA synthesizer. The 42-mer DNA duplex substrate was radiolabeled on its 5′-OH group by treatment with [γ-^{32}P]ATP and T4 polynucleotide kinase. Salts and unincorporated nucleotides were removed by chromatography on Sephadex G50 columns (Pharmacia; Uppsula, Sweden).

Mutagenesis of the I-PpoI gene. Mutations were introduced into the I-*Ppo*I gene in the plasmid pI-Ppo by site-directed oligonucleotide-mediated mutagenesis (Kunkel *et al.*, 1987). This plasmid directs the production of I-*Ppo*I under the control of the T7 promoter (Mannino, S. J. and Raines, R. T., unpublished results). The oligonucleotide sequences used for mutagenesis are shown in Table 4.2. The sequences of all mutant

I-*Ppo*I genes were verified by DNA sequencing using Sequenase Version 2.0 from United States Biochemicals (Cleveland, OH).

Production and purification of D109A, D140A, and K142A I-PpoI from Escherichia coli. I-*Ppo*I mutants were expressed and purified essentially as described by Mannino and Raines (unpublished results). Briefly, starter cultures from freshly transformed cells of *Escherichia coli* BL21(DE3), harboring the plasmid, pI-Ppo, where the gene encoding I-*Ppo*I contained the related mutation, were used to inoculate a larger culture (500 mL) of TB medium (Ausubel *et al.*, 1989) containing ampicillin (200 μg/mL). The inoculated culture was shaken (250 rpm) at 37 °C until late log phase (A = 2.0 OD at 600 nm), and then was induced to express the cDNA that codes for I-*Ppo*I mutants by the addition of IPTG (0.5 mM). Shaking at 37 °C was continued for 2 h after induction, then the cells were harvested by centrifugation.

The cell pellet was resuspended in 50 mL of 10 mM Tris-HCl buffer, pH 7.4, containing EDTA (1 mM) and NaCl (0.4 M), and lysed by passing twice through a French Pressure Cell from SLM Aminco (Urbana, IL). The insoluble debris was removed by centrifugation and washed twice with 50 mL of 50 mM Tris-HCl buffer, pH 8, containing glycerol (5% v/v), EDTA (0.1 mM), and sodium deoxycholate (2% w/v). The resulting pellet was solublized by stirring for 1 h in 10 mL of 10 mM Tris-HCl, pH 8, containing EDTA (1 mM), DTT (1 mM), and guanidine-HCl (6 M). To renature the protein, the solution was diluted over the course of 4 h by the gradual addition of 250 mL of refolding buffer, which was 25 mM CHES/25 mM CAPS buffer, pH 9, containing NaCl (0.10 M), DTT (1 mM), and $MgCl_2$ (10 mM). The suspension was then dialyzed overnight against 20 mM Tris-HCl buffer, pH 9, containing EDTA (1 mM) and, DTT (1 mM). The insoluble material that accumulated during dialysis was removed by centrifugation.

Properly folded I-*Ppo*I variants were purified using an FPLC apparatus from Pharmacia (Uppsala, Sweden). The soluble fraction after dialysis was loaded onto a MonoQ anion exchange column and eluted with a linear gradient of NaCl (0.0 - 0.5 M), in 20 mM Tris-HCl buffer, pH 9, containing EDTA (1 mM). The fractions containing the I-*Ppo*I variant were pooled, dialyzed overnight against 20 mM Tris-HCl, pH 9, EDTA (1 mM), DTT (1 mM), and reinjected onto the anion exchange column. The fractions containing the I-*Ppo*I variant were pooled and purified further by chromatography on a HiLoad™ 26/60 Superdex™ 75 gel filtration column in 50 mM Tris-HCl buffer, pH 7.3, containing EDTA (1 mM), DTT (1 mM), and NaCl (0.25 mM). Purified samples eluted as single symmetric peaks off the gel filtration column as monitored by UV absorbance at 280 nm and migrated as single bands during SDS-PAGE as visualized by Coomassie brilliant blue staining (Figure 4.2). In a typical preparation, 10 mg of enzyme were purified from 250 mL of culture.

Endonuclease cleavage assays. Two substrates containing a single recognition site of I-*Ppo*I, a plasmid and a synthetic oligonucleotide duplex, were used for cleavage assays for wild-type and mutant I-*Ppo*I's. The plasmid p42 contains the recognition site of I-*Ppo*I inserted into the multiple cloning site of pBluescript(KS+) (Muscarella *et al.*, 1990). The synthetic oligonucleotide duplex is a 42-mer and has the sequence, 5′–AAT TAGTAACTATGA**CTCTCTTAAGGTAGC**CAAATGCCTGCA–3′. Reaction mixtures (15 μL) with p42 as the substrate (700 ng) were incubated with increasing concentrations of I-*Ppo*I or I-*Ppo*I mutants (1 pM, 10 pM, 100 pM, 1 nM, or 10 nM of dimeric enzyme) in a 25 mM CAPS/25 mM CHES buffer, pH 10, containing $MgCl_2$ (100 μM and 200 μM) and DTT (1 mM) for 1 h at 37°C. Reactions were quenched with a solution of SDS (1% w/v), EDTA (50 mM), glycerol (10% v/v), and bromophenol blue (0.01% w/v). Products were separated on a 0.75% TAE agarose gel and visualized by staining with ethidium bromide. Similarly, reactions (25 μL) with the ^{32}P-labeled

oligonucleotide substrate (1 μM) were incubated with increasing concentrations of wild-type or mutant I-*Ppo*I (1 pM, 10 pM, 100 pM, 1 nM, or 10 nM of dimeric enzyme) in a 25 mM CAPS/25 mM CHES buffer, pH 10, containing $MgCl_2$ (100 μM), and DTT (1 mM) for 1 h at 37°C. The reactions were quenched with a solution of SDS (1% w/v), EDTA (50 mM), glycerol (10% v/v), and bromophenol blue (0.01% w/v). The products were separated on a 15% TBE polyacrylamide gel, dried, and quantitated using a phosphoimager from Molecular Dynamics (Sunnyvale, CA).

Gel Mobility Shift Assays Gel mobility shift assays were based on the retardation of the electrophoretic mobility of a ^{32}P-labeled DNA molecule upon the binding of a protein (Fried & Crothers, 1981; Garner & Revzin, 1981). These assays were performed essentially as described (Ausubel *et al.*, 1989). The 42-mer DNA duplex was incubated at 25 °C with wild-type or mutant I-*Ppo*I (100 nM of the dimeric enzyme) in a solution of 25 mM CAPS/25 mM CHES buffer, pH 10, containing EDTA (1 mM), DTT (1 mM), glycerol (10% v/v), and bovine serum albumin (50 μg/ml). Mg^{2+} was omitted to prevent DNA cleavage. After 1 h, the binding reactions were loaded directly onto a polyacrylamide (4% w/v) gel running at 30 V/cm. The gels were dried, and ^{32}P was quantitated using a phosphoimager from Molecular Dynamics (Sunnyvale, CA).

4.3 Results

The reactions catalyzed by both I-*Ppo*I and most restriction endonucleases involve the site-specific cleavage of DNA, generating 3′-hydroxyl and 5′-phosphoryl termini, and require only Mg^{2+} ion as a cofactor (Muscarella *et al.*, 1990; Roberts & Halford, 1993). In addition, the primary structure of these proteins contain similar putative active-site motifs, P-(D/E)-X_n-(D/E)-X-K (Table 4.1). Inspired by this similarity, the roles of the corresponding amino acids of I-*Ppo*I, Asp109, Asp140, and Lys142, were explored

through site-directed mutagenesis. Alanine was substituted for these amino acids, and the resulting proteins were tested for DNA cleavage and binding activity.

Catalysis by the I-PpoI mutants. The DNA cleavage activities of wild-type and mutant I-*Ppo*I were measured using two different substrates: the plasmid p42, and a 42-mer synthetic oligonucleotide duplex (Figure 4.3). Assays were performed under optimal conditions for the wild-type enzyme: pH 10 with no exogenous NaCl (Lowery *et al.*, 1992). The activities of the mutant forms of I-*Ppo*I relative to that of wild-type are reported in Table 4.3 and Table 4.4. The cleavage activity of K142A I-*Ppo*I was not detectable with either substrate; thus the relative activity of K142A I-*Ppo*I is decreased by at least 5 orders of magnitude. Lys142 is, therefore, an essential residue for catalysis by I-*Ppo*I. The two aspartate to alanine mutations have only a nominal effect on cleavage. The cleavage activities of D109A and D140A I-*Ppo*I with the plasmid substrate are lowered by only 2- and 5-fold, respectively, relative to that of wild-type I-*Ppo*I. With the 42-mer as the substrate, the cleavage activity decreases 10-fold relative to that of wild-type for both aspartate mutants. These results suggest that the presence or absence of these two aspartate residues has only a small effect on catalysis.

Binding by the I-PpoI mutants. All three mutants bind to DNA with a detectable affinity, as demonstrated by gel mobility shift assays (Figure 4.4). In the absence of Mg^{2+}, D109A, D140A, and K142A I-*Ppo*I exhibit a somewhat lower affinity for the ligand than does the wild-type enzyme. Still, the apparent binding affinities of these mutant proteins are judged to be within 10-fold of that of wild-type. The complexes formed are specific complexes, as unlabeled 42-mer, but not poly(dIdC)·poly(dIdC) (0.2 μg/ml), competed with [^{32}P]42-mer for complex formation (data not shown). In addition, the wild-type protein forms a non-specific complex with slower mobility under these conditions. This complex is not apparent with D109A, D140A, and K142A I-*Ppo*I.

These results indicate that D109A, D140A, and K142A I-*Ppo*I recognize and bind the recognition site.

CD spectroscopy. To evaluate the structural integrity of the mutant proteins, circular dichorism spectra were obtained for wild-type, D109A, D140A, and K142A I-*Ppo*I. The spectra of the mutant proteins did not differ significantly from that of wild-type I-*Ppo*I (Figure 4.5). This result indicates that the mutant proteins have similar secondary structure to that of the wild-type I-*Ppo*I. Using eq. 4.1, the percent α-helix can be estimated (Greenfield & Fasman, 1969).

$$\%\alpha\text{-helix} = \frac{[\theta_{208\ \mathrm{nm}}] - 4{,}000}{33{,}000 - 4{,}000} \tag{4.1}$$

In eq 4.1, θ is the ellipticity at 208 nm in deg•cm^2 dmole^{-1}. This approximation estimates 30% α-helical content in wild-type, D109A, D140A, and K142A I-*Ppo*I.

4.4 Discussion

The amino acid sequence of I-*Ppo*I does not contain the common amino acid motif LAGLI-DADG, of other eukaryotic homing endonucleases (Mueller *et al.*, 1993). The sequence, however, does contain the putative active-site motif of several type II restriction endonucleases (Table 4.1). This motif, P-(D/E)-X_n-(D/E)-X-K, generally consists of two acidic residues and a single basic residue (Thielking *et al.*, 1991). For I-*Ppo*I, this motif is located within the second half of polypeptide chain:

```
M                                P  D        D   K                V
|                                |  |        |   |                |
---------------------------------------------------------------------
1                               108 109     140 142              185
```

where the line represents the polypeptide chain of I-*Ppo*I.

Although little is known about catalysis by homing endonucleases, considerably more is known about catalysis by restriction endonucleases. Analysis of the stereochemistry about the scissile phosphate shows inversion of configuration for both *Eco*RI and *Eco*RV

endonucleases (Connolly *et al.*, 1984; Grasby & Connolly, 1992). Thus, the simplest and generally favored mechanism for catalysis by these enzymes is direct attack by an activated water molecule in-line with the O3′ leaving group (Figure 4.7) (Heitman, 1992; Jeltsch *et al.*, 1992). In this proposed mechanism, a general base deprotonates the attacking water molecule, a positively charged group acts to stabilize the negatively charged pentavalent transition state, and a general acid protonates the leaving group. Efforts have been made to assign these roles to the identified amino acids of the active-site motif and the Mg^{2+} ion cofactor.

The role of the basic residue. The properties of K142A I-*Ppo*I indicate that Lys142 is important for catalysis by I-*Ppo*I. Replacing Lys142 with an alanine residue lowers the catalytic activity of I-*Ppo*I by at least five orders of magnitude (Table 4.3 and Table 4.4), consistent with Lys142 being an active-site residue. The magnitude of decrease in activity is similar to that observed for mutations of the corresponding residue of other endonucleases. For example, the substitution of Lys113 of *Eco*RI by an alanine, a glutamate, a histidine, or a leucine residue results in a decrease in specific cleavage activity of more than five orders of magnitude (Grabowski *et al.*, 1995). The substitution with the positively charged and somewhat isosteric arginine results in only a 10-fold drop in specific cleavage activity (Grabowski *et al.*, 1995). In *Eco*RV, alanine and glutamine substitutions at this position (Lys92) result in a decrease of cleavage activity of at least four orders of magnitude, whereas the glutamate substitution decreases the cleavage activity by three orders of magnitude (Selent *et al.*, 1992). The substitution of the analogous position in *Bam*HI, Glu113 (wild-type) to Lys113(variant), decreases the cleavage activity more than three orders of magnitude and also increases nonspecific binding by the enzyme (Xu & Schildkraut, 1991).

K142A I-*Ppo*I still recognizes and binds its substrate, albeit with a lower apparent affinity (Figure 4.4). This result suggests that the residue plays only a small role in

substrate binding. Evidently, the Lys142 to alanine mutation exerts its greatest effect during catalysis by I-*Ppo*I (Table 4.3 and Table 4.4). The impact of this mutation on catalysis appears to affect k_{cat} as assessments of catalysis were performed in the presence of a substrate concentration (1 μM oligonucleotide) that saturates the wild-type and (presumably) mutant enzymes (Wittmayer & Raines, 1996). This information is consistent with Lys142 having a role in stabilizing the chemical transition-state during catalysis by I-*Ppo*I.

In restriction endonucleases, the role of the lysine residue in the active-site motif remains unclear, but two chemical functions have been considered. First, the positive charge may stabilize the negatively charged pentavalent transition state. Similar roles are observed for basic residues in other nucleases such as Lys41 of ribonuclease A (Trautwein *et al.*, 1991; Messmore *et al.*, 1995), Arg77 of ribonuclease T1 (Heinemann & Saegner, 1982), and Arg87 and Arg35 of staphylococcal nuclease (Cotton *et al.*, 1979). For the endonuclease *Bam*HI, the analogous residue is the acidic Glu113 (Table 4.1). This negatively charged residue may serve a similar function by binding the Mg^{2+} cofactor. In the structure of the *Bam*HI•DNA complex, this alternate glutamic acid creates a potential second metal-ion binding site (Newman *et al.*, 1995). This type of mechanism for *Bam*HI would be similar to the two metal-ion mechanism proposed for the 3′ – 5′ exonuclease activity of *Escherichia coli* DNA polymerase I (Beese & Steitz, 1991).

A second role considered for the active-site basic residue of restriction endonucleases is to act as the general base that deprotonates the attacking water molecule (Figure 4.7) (Kostrewa & Winkler, 1995). From the structure of the *Eco*RV – substrate complex, this role is plausible. The basic residue Lys92 forms a hydrogen bond to a water molecule, which is located properly for in-line attack of the scissile phosphate. Chemically, however, this role is less plausible as there appears to be little reason to expect a lowered

pK_a of the *Eco*RV lysine residue (Kostrewa & Winkler, 1995). In the amino acid active-site motif of *Bam*HI, the corresponding residue is not a lysine but the acidic residue Glu113. Still, the precise role of Lys142 in catalysis by I-*Ppo*I remains to be determined.

The specific binding of K142A I-*Ppo*I to DNA suggests that the mutant protein is properly folded (Figure 4.4). Both the circular dichroism spectra and gel-filtration elution volume of K142A I-*Ppo*I are parallel to those of the wild-type enzyme, suggesting that no major conformational changes have occurred in the secondary structure of the protein dimer (Figure 4.5 and Figure 4.6). Still, without detailed structural information, the possibility that the Lys142 to alanine mutation altered the tertiary structure of the endonuclease rendering it catalyticaly inert and that the lysine residue is remote from the active site can not be eliminated.

The role of the acidic residues. The properties of D109A and D140A I-*Ppo*I suggest that these residues are not essential residues for catalysis by I-*Ppo*I. The relative cleavage and binding activity of these alanine mutants decrease no more than 10-fold as compared to that of the wild-type enzyme. These results imply that Asp109 and Asp140 are not part of the active site of I-*Ppo*I. Substitutions at this position in other endonucleases show much greater decreases in cleavage activity. For example, substituting an alanine or asparagine residue for Asp74 or Asp90 of *Eco*RV results in a decrease in cleavage activity by more than four orders of magnitude (Selent *et al.*, 1992). In addition, these mutations cause *Eco*RV to bind DNA more strongly (Selent *et al.*, 1992). The substitution of a glutamate residue for Asp74 results in a 100-fold decrease in cleavage activity; the same substitution for Asp90 results in no change in activity (Selent *et al.*, 1992). In *Eco*RI, substitution of an alanine or asparagine for Asp91 shows a decrease in activity that is greater than five orders of magnitude. The same mutations at the second acidic position, Glu111, showed a decrease in activity of 100-fold (Grabowski *et al.*, 1995). Similar results have been obtained with mutation of the acidic residues of

the active-site motifs of *Bam*HI, *Fok*I, and *Nae*I endonucleases (Xu & Schildkraut, 1991; Waugh & Sauer, 1993; Holtz & Topal, 1994).

In restriction endonucleases, the proposed role of the acidic residues is to chelate the Mg^{2+} ion cofactor. Crystals of the *Eco*RV•DNA complex, soaked in a solution containing Mg^{2+} ions, show that the cofactor is coordinated by the side chains of Asp74 and Asp90, and a phosphoryl oxygen of the scissile phosphate (Figure 4.1) (Kostrewa & Winkler, 1995). The results presented for I-*Ppo*I containing alanine mutations at the corresponding positions, Asp109 and Asp140, suggest that these residues are not required for Mg^{2+} binding. Only a small decrease in catalytic activity was observed even at low concentrations of Mg^{2+}, suggesting that the mutations did not disrupt the affinity of the endonuclease for the metal-ion cofactor. These residues, therefore, do not appear to function analogously to the acidic residues of the restriction endonucleases.

The discovery of two putative homing endonucleases, I-*Dir*I and I-*Naa*I, within the nuclear introns of the myxomycete, *Didymium iridis*, and the amoeba-flagellate, *Naegleria andersonia,* revealed a conserved region with the amino acid sequence of I-*Ppo*I (Johansen *et al.*, 1993). This region of 30 amino acids contains three cysteine and two histidine residues (Figure 4.8). Called the His-Cys box, this region resembles the sequences of zinc fingers and was recently proposed to compose a metal-ion binding site (Belfort & Perlman, 1995). Perhaps this region of I-*Ppo*I coordinates the metal-ion cofactor rather than Asp109 and Asp140. Also consistent with this proposal is our finding that the metal-ion binding site of I-*Ppo*I, which accepts Mg^{2+}, Mn^{2+}, Ca^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+} as a cofactor (Wittmayer & Raines, 1996), has a more relaxed specificity than does that of the restriction endonucleases.

Mutational analysis of the P-(D/E)-X_n-(D/E)-X-K motif apparent in the primary structure of the homing endonuclease, I-*Ppo*I, argues that this motif does not describe the active site of I-*Ppo*I. The basic residue of the motif, Lys142, is essential for catalysis by

I-*Ppo*I; however, the two acidic residues, Asp109 and Asp140, do not appear to be essential catalytic residues. Despite the similarities of the reactions catalyzed by both homing endonucleases and restriction endonucleases, the active-site of I-*Ppo*I appears to employ a distinct architecture to catalyze the site-specific cleavage of DNA.

Table 4.1: Putative Active-Site Motif[a]

I-*Ppo*I	$P_{108}D_{109}$	···	$D_{140}N_{141}K_{142}$
*Eco*RI	P_{90} D_{91}	···	$E_{111}A_{112}K_{113}$
*Eco*RV	P_{73} D_{74}	···	D_{90} I_{91} K_{92}
*Bam*HI	D_{94}	···	$E_{111}F_{112}E_{113}$
*Pvu*II	D_{58}	···	E_{68} L_{69} K_{70}
*Fok*I	$P_{449}D_{450}$	···	$D_{467}T_{468}K_{469}$
*Nae*I	E_{70}	···	D_{95} C_{96} K_{97}

[a] Based on Thielking *et al.* (1991) and Anderson (1993).

Table 4.2: Oligonucleotides Used for Site-Directed Mutagenesis of I-*Ppo*I Gene

Oligonucleotide	Sequence
D109A	GCCGTTGATGTT**CGC**CGGTTCTAATAG
D140A	TCTGCCTTTGTT**CGC**GTCTAGTGACTC
K142A	CCAGTTTCTGCC**CGC**GTTGTCGTCTAG

[a] The CGC anticodon in bold print indicates the alanine mutation.

Table 4.3: Relative DNA Cleavage Activity of Wild-Type, D109A, D140A, and K142A I-*Ppo*I with Plasmid Substrate

	Relative Activity	
I-*Ppo*I	pH 10, 100 μM $MgCl_2$	pH 10, 200 μM $MgCl_2$
wild-type	1	1
D109A	0.2	0.5
D140A	0.1	0.2
K142A	< 0.00001	< 0.00001

Table 4.4: Relative DNA Cleavage Activity of Wild-Type, D109A, D140A, and K142A I-*Ppo*I with 42-mer as a Substrate

I-*Ppo*I	Relative Activity pH 10, 100 μM $MgCl_2$
WT	1
D109A	0.1
D140A	0.1
K142A	< 0.00001

Figure 4.1 Active Site of Restriction Endonucleases, *Eco*RI and *Eco*RV. Structures are based on the coordinates of the *Eco*RI•DNA and *Eco*RV•DNA complexes (Brookhaven data bank entries, 1R1E and 3RVE) (Frederick *et al.*, 1984; Kim *et al.*, 1990; Kostrewa & Winkler, 1995). For clarity, only the amino acid residues of the active-site motif of one subunit and the scissile phosphate of one DNA strand are shown. The amino acids are in red, the DNA strands are in green, and the Mg^{2+} ion cofactor is in black as a shaded sphere.

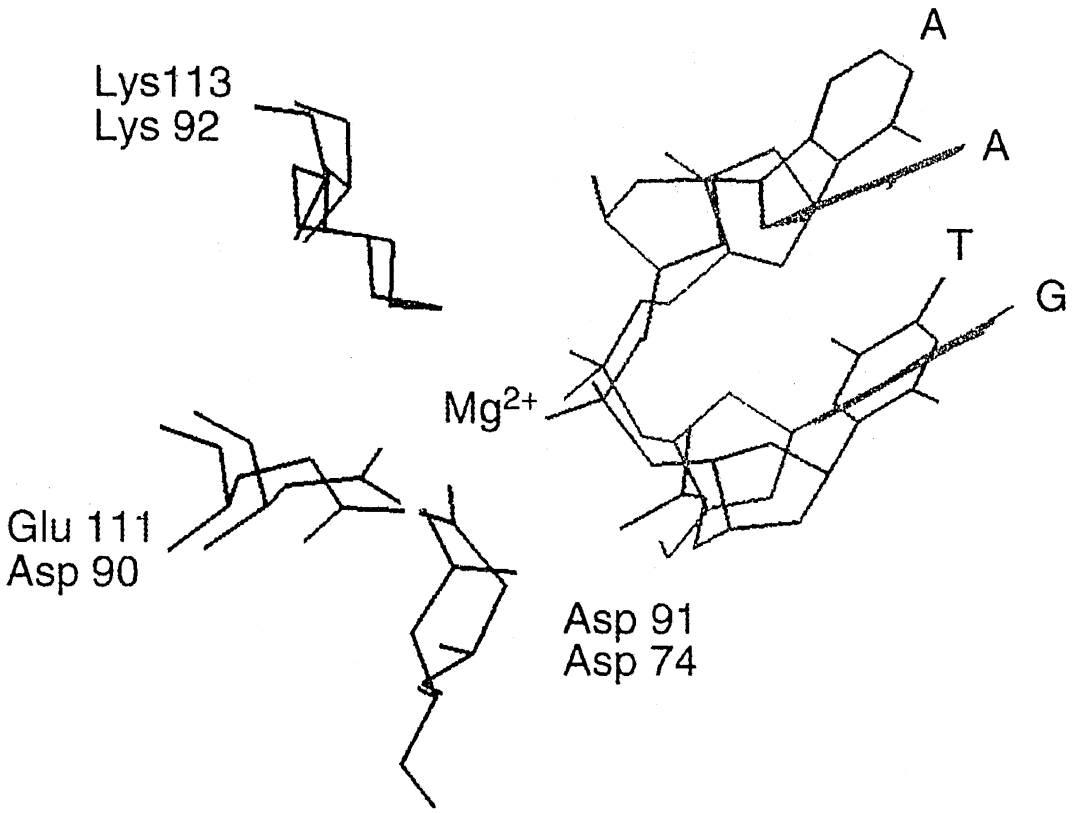
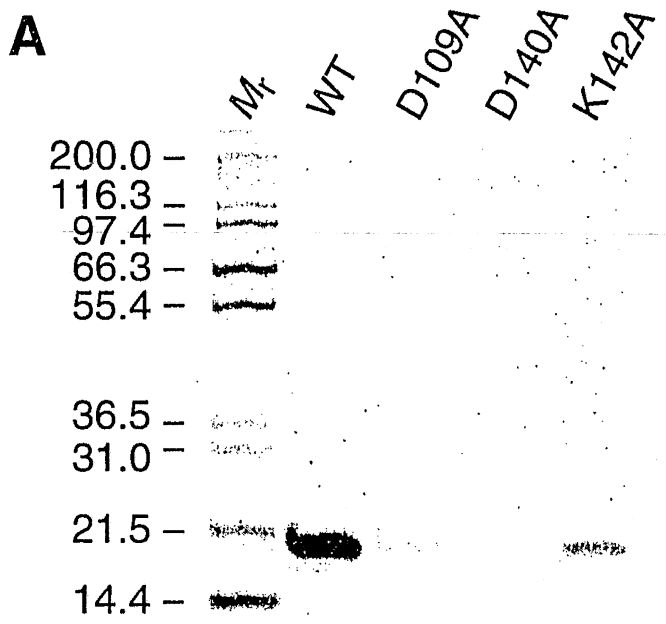

Lys113
Lys 92
A
A
T
G
Mg2+
Glu 111
Asp 90
Asp 91
Asp 74

Figure 4.2 Purity of wild-type and mutant I-*Ppo*I endonucleases. (A) Gel after SDS–PAGE. (B) Gel filtration profile. The proteins were subjected to chromatography on a HiLoad™ 26/60 Superdex™ 75 gel filtration column in 50 mM Tris-HCl buffer, pH 7.3, containing EDTA (1 mM), DTT (1 mM), and NaCl (0.25 mM).


A
Mr
WT
D109A
D140A
K142A
200.0
116.3
97.4
66.3
55.4
36.5
31.0
21.5
14.4

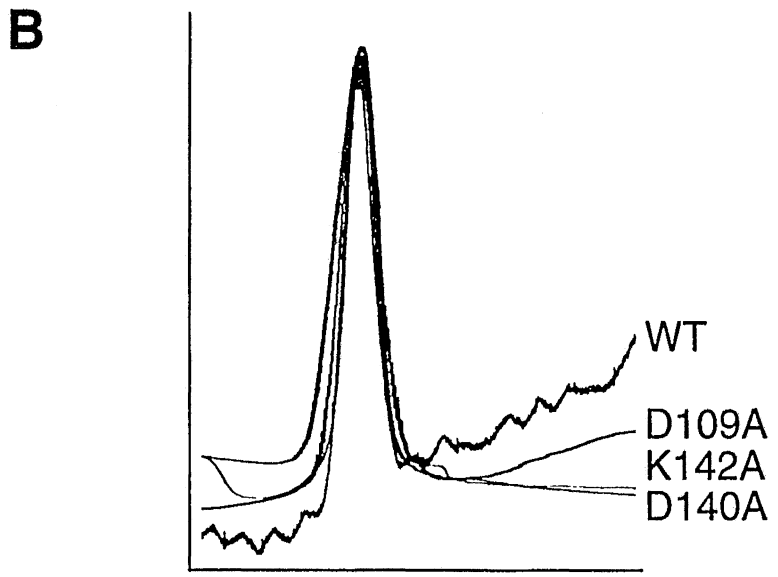

B
WT
D109A
K142A
D140A

Figure 4.3 Cleavage activity of wild-type and mutant I-*Ppo*I. The plasmid substrate was incubated with increasing concentrations of I-*Ppo*I or I-*Ppo*I mutants (1 pM, 10 pM, 100 pM, 1 nM, or 10 nM of dimeric enzyme) in a 25 mM CAPS/ 25 mM CHES buffer, pH 10, containing $MgCl_2$ (100 μM) and DTT (1 mM) for 1 h at 37 °C. Reactions were quenched with a solution of SDS (1% w/v), EDTA (50 mM), glycerol (10% v/v), and bromophenol blue (0.01% w/v). Products were separated on an agarose (0.75% w/v) gel in TAE buffer and visualized by staining with ethidium bromide. The lanes represent increasing concentration of enzyme: lane 1, 0 pM enzyme; lane 2, 1 pM; lane 3, 10 pM; lane 4, 100 pM; lane 5, 1 nM; lane 6, 10 nM.

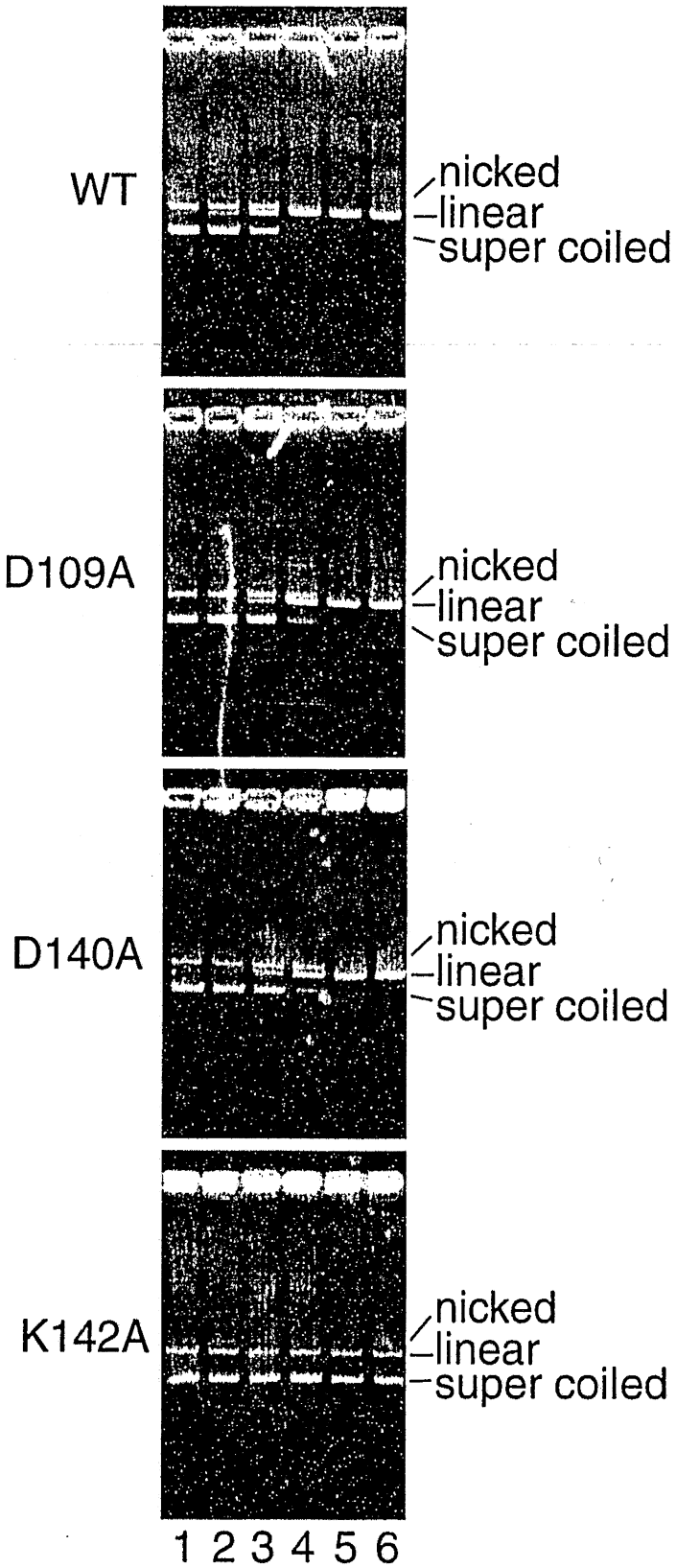

WT
nicked
linear
super coiled
D109A
nicked
linear
super coiled
D140A
nicked
linear
super coiled
K142A
nicked
linear
super coiled
1 2 3 4 5 6

Figure 4.4 DNA binding by wild-type and mutant I-*Ppo*I as observed by gel mobility shift assays. The ^{32}P-labeled 42-mer (20 nM) was incubated with 100 nM of protein in 25 mM CAPS/25 mM CHES buffer, pH 10, containing EDTA (1 mM), DTT (1 mM), glycerol (10% v/v), and bovine serum albumin (50 μg/mL). Mg^{2+} was omitted to prevent DNA cleavage.

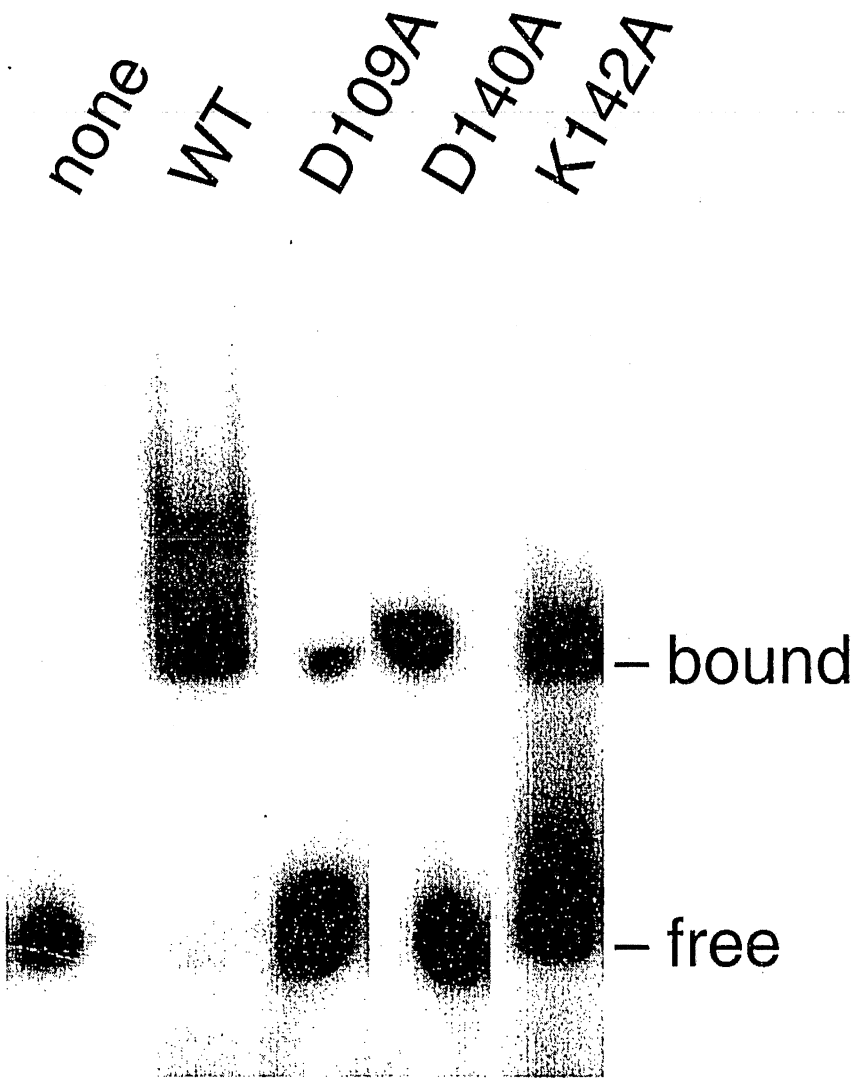

none
WT
D109A
D140A
K142A
– bound
– free

Figure 4.5 Circular dichroism spectra of wild-type and K142A I-*Ppo*I. Spectra were measured in a 1 mm pathlength cell for proteins (250 μg/mL) in 25 mM CAPS/25 mM CHES buffer, pH 10, containing EDTA (1 mM).

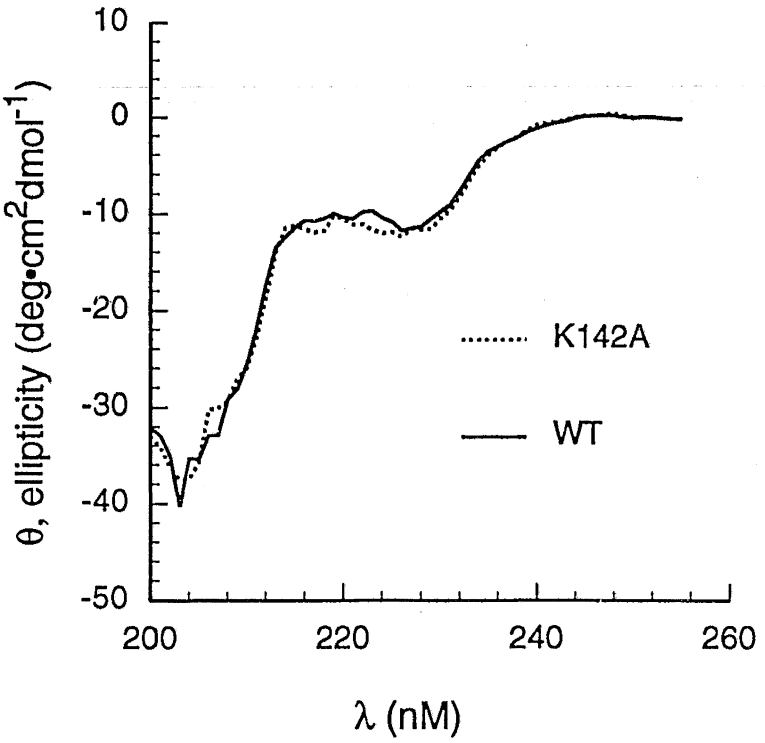

10
0
-10
-20
-30
-40
-50
θ, ellipticity (deg•cm²dmol⁻¹)
200
220
240
260
λ (nM)
K142A
WT

Figure 4.6 Gel filtration profile of K142A I-*Ppo*I at pH 10 (reactions conditions). The purified protein was subjected to chromatography on a HiLoad™ 26/60 Superdex™ 75 gel filtration column using 25 mM CAPS/25 mM CHES buffer, pH 10, containing EDTA (1 mM). The protein eluted as a symmetric peak similar to that of the dimeric wild-type protein.

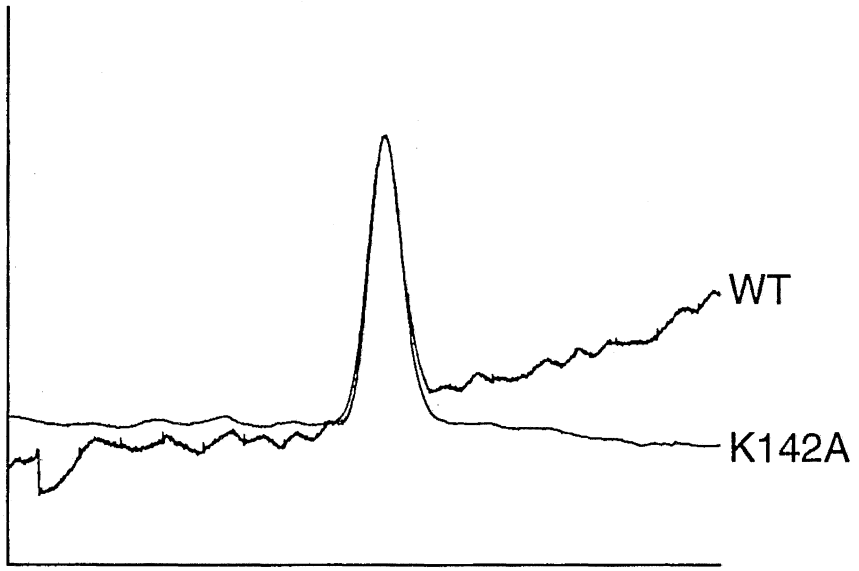

WT
K142A

Figure 4.7 Mechanism of Heitman (1992) for DNA cleavage by restriction endonuclease *Eco*RI. Direct in-line attack of an activated water molecule on the P–O3′ bond. In this mechanism, a general base deprotonates the attacking water molecule, a positively charged group acts to stabilize the negatively charged pentavalent transition state, and a general acid protonates the leaving group. These roles could be fulfilled by functional groups on the enzyme and the Mg^{2+} ion cofactor.

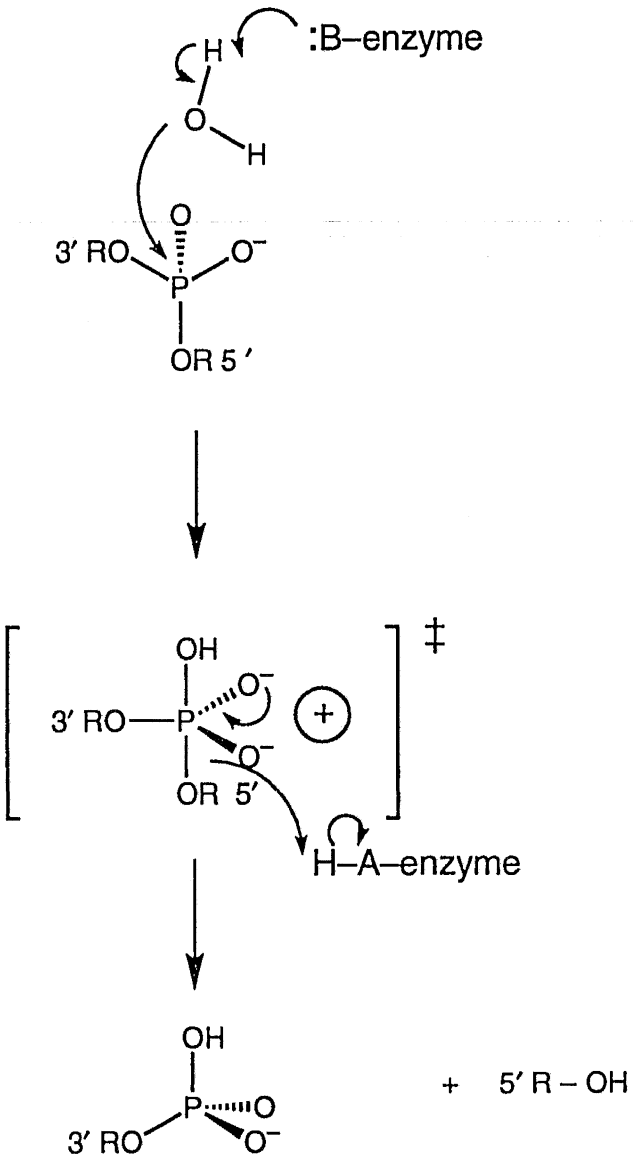

:B–enzyme
H
O
H
O
3′ RO
O⁻
P
OR 5′
OH
O⁻
3′ RO
P
O⁻
OR 5′
+
‡
H–A–enzyme
OH
P
O
3′ RO
O⁻
+ 5′ R – OH

Figure 4.8 The His-Cys box. Conserved amino acids between I-*Ppo*I and two putative homing endonucleases, I-*Dir*I and I-*Naa*I (Johansen *et al.*, 1993). Conserved residues are boxed, and the conserved His and Cys residues are denoted with an asterisk.

```
                    *   *           *           *                                   *
I-Ppo I  116  -  T  A  S H L C  -  H N T R C H N P L H  -  L C W E S L D D N K G R N W C P  149
I-Naa I  176  -  T  I  S H L C  -  G N G G C A R P G H  -  L R I E K K T V N D E R T H C H  210
I-Dir I  169  -  H  S  S H L C  K  G D G S C M E L K H  T  L R V P A Q T N L A D H E L C P  204
```

Appendix

Supplementary Data for Chapter 3

Appendix: Effect of Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I: All Sequences Assayed

Identification	Sequence[a]	Activity Class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
1	**C**TGA CTCTCTTAAGGTAG**T** C**T**AA	1
2	**G**TG**T** CTCTCTT**C**AGGTAGC C**TT**A	1
3	ATGA CTCTCTTAAGGTAGC CAAA	1
4	**G**TGA CTCTCTTAAGGTAGC **A**AA**G**	3
5	ATGA CTCTCTTAAGGTAGC C**G**AA	1
6	ATGA CTCTCTTAA(**T**)GGTAGC C**T**AAA	4
7	ATGA CTCTCTTAAGGTAGC CAAA	1
8	ATGA CTCTCTTAAGGTA**T**C CAAA	2
9	ATGA CTCTCTTAAGG**C**A**T**C CA**T**A	2
10	ATGA CTCTCTTAAGGTAGC CAAA	1
11	ATGA **A**TCT**T**TTAAGGT**T**GC CAAA	4
12	ATGA CTCTCTTAAGGTAG**G** CAAA	2
13	AT**A**A CTCTCTTAAG**T**TA**C**C CAAA	4
14	A**G**GA **T**TCTCTTAAGGTAGC CAAA	4
15	ATG**T** CTCTCTTAAGGTAGC CA**G**A	2
16	ATGA CT**A**TCTTAA**A**GTAGC CAAA	3
17	ATGA C**A**CTCT**AG**AGGTAGC CAAA	4
18	ATGA CTCT**AC**TAAGGTAGC CAA**C**	4
19	**AA**GA CTCTC**CAAC**GGTAGC CAAA	4
20	ATGA CTCTCTTAAGGTAGC CAAA	1
21	ATGA CTCTCTT**C**AGGTAGC CAA**T**	2.5[c]
22	ATGA CTC**A**CTT**T**AGGTAG**A** C**C**AA	4
23	ATGA CTCTC(**T**)TTAAGGTAGC CAAA	4
24	ATGA CTCTCTTAAGGTAGC CAAA	1
25	ATGA CT**G**T**T**TA**A**AG**C**TAGC CAAA	4

Appendix: Effect of Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I: All Sequences Assayed

Identification	Sequence[a]	Activity Class[b]
26	ATGA CTCTCGTAACGTAGC CATA	2.5[c]
27	ATGA CTTTCTTTAGGGAGC CATA	4
28	ATGA CTCACTTAAGGTAGC CAAC	3
29	ATGA CTCTCTTAAGGTAGC CAAA	1
30	ATGA CTCTCCTAAGGTAGC CAAA	3
31	ATGA CTCGCTTAAGGCAGT CAAA	2
32	ATGA CTCTCTTAΔGGTAGC CGAA	4
33	ATGA CTCTCGTAAGGTAGC CATA	1
34	TTGA CTCTCTTAAGGTAGC CTAA	1
35	AGGA CTTTCTAAAGGTAGC CAAA	2
36	ATGA CTCTCTTAATGTAGC CAAA	2
37	ATGA CTCTCTTAAGTTATC CAAA	3
38	ATGA CTCTCTTAAGGTAGC CAAA	1
39	TTGC CTCTCTTAAGGTAGC CAGA	3
40	ATGA CΔCTCTTAAGGTAGC CAAA	4
41	ATGA CTCTCTTGAGGTAGC CAAC	2
42	ATGG CTCTCTTAAGGTAGG CATA	2
43	ATGC CTCTCTTGAGGTAGC CAAC	4
44	ATGA CTCACΔTAAGGTAGC CAAA	4
45	ATGA CTATTTAAAGGTAGC CAAG	2
46	ATGA CGCTCTTTAGGTAGC CAAA	1
47	ATGA CTCTCTTAAGGTAGC CAAA	1
48	AAGA CTATCTTAAGGTAGC CAAC	1
49	ATGT CTCGCTTAAGGTAGC CAAA	3
50	ATTA CACTATTAGGGTAGC CAAA	4

Appendix: Effect of Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I: All Sequences Assayed

Identification	Sequence[a]	Activity Class[b]
51	ATGA CACTCTTAAGGTAGG CAAA	3
52	TTGA CTTTCTTAAGGTAGT CAAA	2.5[c]
53	ATGA CTCTCTTAAGCTCGC CCCA	4
54	ATCA CTCTCTTAAGGTAGC TAAA	2
55	ATGT CTCTCTTAAGGTAGC CAAC	1
56	ATGG CTTTCTTAAGGTAGC CGGA	3
57	ATGA CTCTCTTAAGCTAGC CAAA	2
58	GTGA CTCTCTTAAGCTAGC AAGA	4
59	ATGA CTCACTTTGGGTAGC CAAA	4
60	ATGA CTCTCTTGAGGTAGC CAAA	1
61	ATGA CTCTCGTAACGTAAC CATA	3
62	ATGA CTCTCTTTAGGTAGC CACA	1
63	ATGA CTCTCTTAAGGTAGC CAAA	1
64	ATGA CTCTCTTAAAGTAGC CAGA	3
65	CTGA CTCTCTTCAGGTCGC CAAA	4
66	ATGA CTCACTTAAGGTAGC CAAG	2
67	ATGT CTCTCTTAAGGTAGC CAAA	1
68	ATGA CTCTCTTAAGGTCGC CAAC	4
69	ATGA CTCTCTTAAGGTAGC CAAA	1
70	ATGG CTCTCTTAAGGTTGC CAAA	4
71	AGGA CTCTCTTAAGGTAGC TAAT	2
72	ATGA CTCTCTTAAGGTAGC CAGA	1
73	ATGA CTCTCTAAAGGTATC CAAA	4
74	ATGA CTCTCTTAAGGTCGC CAAA	4
75	ATGA ATCTCTTAAGGTAGC CAAA	2
76	ATGA CCCTCTTGAGGTAGC CCAA	3
77	ATGA CTCTCTAAAGGAAGC CTGT	3.5[c]

Appendix: Effect of Basepair Mutations in the Substrate on Catalysis by I-*Ppo*I: All Sequences Assayed

Identification	Sequence[a]	Activity Class[b]
S1	ATGA **T**TC**A**CTTAAGGTAGC CAAA	4
S2	ATGA **T**TC**A**CTTAAGGTAGC CAAA	4
S3	ATGA **A**TC**A**CTTAAGGTAGC CAAA	4
S4	ATGA **A**TCTCTTAAGGTAGC CAAA	2
S5	ATGA **T**TCTCTTAAGGTAGC CAAA	2.5[c]
S6	ATGA **A**TCTCTTAAGGTAGC CAAA	2
S7	ATGA C**C**CT**A**TTAAGGTAGC CAAA	4
S8	ATGA C**G**CT**A**TTAAGGTAGC CAAA	4
S9	ATGA C**C**CT**A**TTAAGGTAGC CAAA	4
S10	ATGA CTCT**A**TTAAGGTAGC CAAA	4
S11	ATGA C**A**CTCTTAAGGTAGC CAAA	2.5[c]
S12	ATGA CTΔTCTTAAGGTAGC CAAA	4
S13	ATGA CT**T**TC**A**TAAGGTAGC CAAA	3
S14	ATGA CT**T**TCTTAAGGTAGC CAAA	2
S15	ATGA CT**A**TCTTAAGGTAGC CAAA	1
S16	ATGA CT**T**TC**A**TAAGGTAGC CAAA	3
S17	ATGA CTC**G**CTTAAGGTAGC CAAA	1.5[c]
S18	ATGA C**A**CTCTTAAGGTAGC CAAA	2.5[c]
S19	ATGA C**C**CTCTTAAGGTAGC CAAA	3
S20	ATGA CTC**A**CTΔAAGGTAGC CAAA	4
S21	ATGA CTC**G**CT**A**AAGGTAGC CAAA	3
S22	ATGA CTC**A**CΔTAAGGTAGC CAAA	4
S23	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S24	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S25	ATGA CTCT**T**TT**T**AGGTAGC CAAA	3

[a] Substituted bases are highlighted with bold print. Inserted bases are highlighted with bold print in parentheses. Deleted bases are represented by a Δ. The four leftmost and rightmost bases are part of flanking regions.

[b] Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37° in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1 nM, 10 nM, or 100 nM). Class 1 substrates were unaffected; they cleaved as efficiently as wild type. Class 2 were affected slightly; 70–95% of substrate was cleaved at the lowest enzyme:DNA ratio. Class 3 were affected moderately; 100% of the substrate was cleaved only at the highest enzyme:DNA ratio. Class 4 were affected greatly; the substrate was not cleaved completely at even the highest enzyme:DNA ratio.

[c] When the results of two individual assays varied the average of the two assays is given.

References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899-907.

Aiken, C. R., Fisher, E. W. & Gumport, R. I. (1991). The specific binding, bending, and unwinding of DNA by *Rsr*I endonuclease, an isoschizomer of *Eco*RI endonuclease. *J. Biol. Chem.* **266**, 19063-19069.

Anderson, J. E. (1993). Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.* **3**, 24-30.

Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290**, 754-758.

Attie, A. D. & Raines, R. T. (1995). Analysis of receptor – ligand interactions. *J. Chem. Educ.* **72**, 119-124.

Ausubel, F. M., Thomas, P. J., Mills, A., Moss, D. S. & Palmer, R. A. (1989). *Current Protocols in Molecular Biology*, Wiley, New York.

Barik, S. (1993). Site-directed mutagenesis by double polymerase chain reaction–megaprimer method. In *PCR Protocols: Current Methods and Applications* (B. A. White, Ed.), pp. 277-286, Humana Press Inc, Totowa, NJ.

Beese, L. S. & Steitz, T. A. (1991). Structural basis for the 3′ - 5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25-33.

Belfort, M. (1989). Bacteriophage introns: parasites within parasites. *Trends Genet.* **5**, 209-213.

Belfort, M. & Perlman, P. S. (1995). Mechanisms of intron mobility. *J. Biol. Chem.* **270**, 30237-30240.

Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990). 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. *Nucleic Acids Res.* **18**, 3763-3770.

Bell-Pedersen, D., Quirk, S. M., Aubrey, M. & Belfort, M. (1989). A site-specific endonuclease and co-conversion of flanking exons associated with the mobile td intron of phage T4. *Gene* **82**, 119-126.

Biburger, M., Niederweiss, M. & Hillen, W. (1994). Oligo [d(C)•d(G)] runs exhibit a helical repeat of 11.1 bp in solution and cause slight DNA curvature when properly phased. *Nucleic Acids Res.* **22**, 1562-1566.

Blackburn, G. M. & Gait, M. J. (1990). *Nucleic Acids in Chemistry and Biology*, IRL Press, Oxford, New York, Tokyo.

Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D. & Knowles, J. R. (1988). Triosephosphate isomerase catalysis is diffusion controlled. *Biochemistry* **27**, 1158-1167.

Bos, J. L., Heyting, C., Borst, P., Arnberg, A. C. & Van Bruggen, E. F. J. (1978). An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA. *Nature* **275**, 336-338.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Burger, G. & Werner, S. (1985). The mitochondrial *URF1* gene in *Neurospora crassa* has an intron that contains a novel type of URF. *J. Mol. Biol.* **186**, 231-242.

Burke, J. M. (1988). Molecular genetics of group I introns: RNA structures and protein factors required for splicing—a review. *Gene* **73**, 273-294.

Byrk, M., Quirk, S., Mueller, J. E., Loizos, N., Lawrence, C. & Belfort, M. (1993). The td intron endonuclease I-*Tev*I makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J.* **12**, 2141-2149.

Cech, T. R. (1987). The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**, 1532-1539.

Cech, T. R. (1988). Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. *Gene* **73**, 259-271.

Cech, T. R. (1990). Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**, 543-568.

Cech, T. R., Tanner, N. K., Tinoco, I. J., Weir, B. R., Zuker, M. & Perlman, P. S. (1983). Secondary structure of the *Tetrahymena* ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. *Proc. Natl. Acad. Sci. USA* **80**, 3903-3907.

Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J. E. (1994). Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* **13**, 3927-3935.

Chu, F. K., Maley, F., Wang, A.-M., Pedersen-Lane, J. & Maley, G. (1991). Purification and substrate specificity of a T4 phage intron-encoded endonuclease. *Nucleic Acids Res.* **19**, 6863-6869.

Cleland, W. W. (1979). Statistical analysis of enzyme kinetic data. *Methods Enzymol.* **63**, 103-138.

Coen, D., Deutsch, J., Netter, P., Petrochilo, E. & Slonimski, P. P. (1970). *Mitochondrial genetics, I. Methodology and phenomenology*. **24**, 449-496 (Cambridge University Press, Cambridge).

Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986). 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.

Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. (1988). Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc. Natl. Acad. Sci. USA* **85**, 6022-6026.

Colleaux, L., Micheal-Wolwertz, M.-R., Matagne, R. F. & Dujon, B. (1990). The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. *Mol. Gen. Genet.* **223**, 288-296.

Connolly, B. A., Eckstein, F. & Pingoud, A. (1984). The stereochemical course of the restriction endonuclease *Eco*RI catalyzed reaction. *J. Biol. Chem.* **259**, 10727-10733.

Cooper, A. A., Chen, Y.-J., Lindorfer, M. A. & Stevens, T. H. (1993). Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J.* **12**, 2575-2583.

Cooper, A. A. & Stevens, T. H. (1995). Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci.* **20**, 351-356.

Cote, V., Mercier, C., Lemieux, C. & Turmel, M. (1993). The single group-I intron in the chloroplast *rrn*L gene of *Chlamydomonas humicola* encodes a site-specific DNA endonuclease. *Gene* **129**, 69-79.

Cotton, F. A., Hazen, E. E. & Legg, M. J. (1979). Staphylococcal nuclease: proposed mechanism of action based on structure of enzyme–thymidine 3′,5′-bisphosphate-calcium ion complex at 1.5 Å resolution. *Proc. Natl. Acad. Sci. USA* **76**, 2551.

Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1993). A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. USA* **90**, 5414-5417.

Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1994). Purification and characterization of two forms of I-*Dmo*I, a thermophilic site-specific endonuclease encoded by an archaeal intron. *J. Biol. Chem.* **269**, 28885-28892.

Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300**, 719-724.

Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J. & Sedgwick, S. G. (1992). Protein splicing in the maturation of *M. tuberculousis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* **71**, 201-210.

Delahodde, A., Goguel, V., Becam, A. M., Creusot, F., Perea, J., Banroques, J. & Jacq, C. (1989). Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell* **56**, 431-441.

Doolittle, R. F. (1993). The comings and goings of homing endonucleases and mobile introns. *Proc. Natl. Acad. Sci. USA* **90**, 5379-5381.

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

Dujon, B. (1981). *Mitochondrial genetics and functions*, Cold Spring Harbor Press, Cold Spring Harbor, NY.

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

Dujon, B., Belfort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P. S. & Vogt, V. M. (1989). Mobile introns: definition of terms and recommended nomenclature. *Gene* **82**, 115-118.

Dujon, B., Cottarel, G., Colleaux, L., Betermier, M., Jacquier, A., d'Auriol, L. & Gailbert, F. (1985). Mechanism of integration of an intron within a mitochondrial gene: a double strand break and the transposase function of an intron encoded protein as revealed by in vivo and in vitro assays. In *Achievements and Perspectives of Mitochondrial Research* (E. Quagliariello, E. C. Slater, F. Palmieri, C. Saccone & A. M. Kroon, Ed.), pp. 215-225, Elsevier, Amsterdam.

Dürrenberger, F. & Rochaix, J.-D. (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. & Rochaix, J.-D. (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.

Eddy, S. R. & Gold, L. (1991). The phage T4 *nrd*B intron: a deletion mutant of a version found in the wild. *Genes Dev.* **5**, 1032-1041.

Ellison, E. L. & Vogt, V. M. (1993). Interaction of the intron-encoded mobility endonuclease I-*Ppo*I with its target site. *Mol. Cell Biol.* **13**, 7531-7539.

Faye, G., Dennebouy, N., Kujawa, C. & C., J. (1979). Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast Saccharomyces cerevisiae. *Mol. Gen. Genet.* **168**, 101-109.

Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B. C., Greene, P., Boyer, H. W. & Rosenberg, J. M. (1984). Kinked DNA in crystalline complex with *Eco*RI endonuclease. *Nature* **309**, 327-331.

Fried, M. & Crothers, D. M. (1981). Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505-6525.

Garner, M. M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047-3060.

Gauthier, A., Turmel, M. & Lemieux, C. (1991). A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. *Curr. Genet.* **19**, 43-47.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.

Gimble, F. S. & Stephens, B. W. (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.

Gimble, F. S. & Thorner, J. (1992). Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae. Nature* **357**, 301-306.

Goodrich-Blair, H., Scariato, V., Gott, J. M., Xu, M.-Q. & Shub, D. A. (1990). A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. *Cell* **63**, 417-424.

Goodrich-Blair, H. & Shub, D. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**, 211-221.

Gorbalenya, A. (1994). Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci.* **3**, 1117-1120.

Gott, J. M., Zeech, A., Bell-Pederson, D., Ehrenman, K., Belfort, M. & Shub, D. A. (1988). Genes within genes: Independent expression of phage T4 intron open reading frames and the genes in which they reside. *Genes Dev.* **2**, 1791-1799.

Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G. & Alves, J. (1995). Site-directed mutagenesis in the catalytic center of the restriction endonuclease *Eco*RI. *Gene* **157**, 113-118.

Grasby, J. A. & Connolly, B. A. (1992). Stereochemical outcome of the hydrolysis reaction catalyzed by the *Eco*RV restriction endonuclease. *Biochemistry* **31**, 7855-7861.

Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E. & Rosenberg, J. M. (1981). Sequence analysis of the DNA encoding the *Eco*RI endonuclease and methylase. *J. Biol. Chem.* **256**, 2143-2153.

Greenfield, N. & Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108-4116.

Halford, S. E. & Goodall, A. J. (1988). Modes of DNA cleavage by the *Eco*RV restriction endonuclease. *Biochemistry* **27**, 1771-1777.

Halford, S. E. & Johnson, N. P. (1980). The *Eco*RI restriction endonuclease with bacteriophage lambda DNA. Equilibrium binding studies. *Biochem. J.* **191**, 593-604.

Halford, S. E. & Johnson, N. P. (1981). The *Eco*RI restriction endonuclease, covalently closed DNA and ethidium bromide. *Biochem. J.* **199**, 767-777.

Heinemann, U. & Saenger, W. (1982). Specific protein-nucleic acid recognition in ribonuclease T1-2′-guanylic acid complex : an X-ray study. *Nature* **299**, 27-31.

Heinemann, U., Alings, C. & Hahn, M. (1994). Crystallographic studies of the DNA helix structure. *Biopys. Chem.* **50**, 157-167.

Heitman, J. (1992). How the *Eco*RI endonuclease recognizes and cleaves DNA. *BioEssays* **14**, 445-454.

Henke, R. M., Butow, R. A. & Perlman, P. S. (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., Bonen, L., da Haan, M., van der Horst, G. & Grivell, L. A. (1983). Two intron sequences in yeast mitochondrial *COXI* gene: Homology among URF-containing introns and strain dependent variation in flanking exons. *Cell* **32**, 379-389.

Hermes, J. D., Blacklow, S. C. & Knowles, J. R. (1990). Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc. Natl. Acad. Sci. USA* **87**, 696-700.

Hodges, R. A., Perler, F. B., Noren, C. J. & Jack, W. E. (1992). Protein splicing removes intervening sequences in an archaea DNA polymerase. *Nucleic Acids Res.* **20**, 6153-6157.

Holtz, J. K. & Topal, M. D. (1994). Location of putative binding and catalysis sites of *Nae*I by random mutagenesis. *J. Biol. Chem.* **269**, 27286-27290.

Hyde-DeRuyscher, R. P., Jennings, E. & Shenk, T. (1995). DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res.* **23**, 4457-4465.

Inoue, T. & Cech, T. R. (1985). Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **82**, 648-652.

Jaquier, A. & Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383-394.

Jeltsch, A., Alves, J., Maass, G. & Pingoud, A. (1992). On the catalytic mechanism of *Eco*RI and *Eco*RV. A detailed proposal on biochemical results, structural data and molecular modeling. *FEBS Lett.* **304**, 4-8.

Jen-Jacobson, L., Kurpiewski, M., Lesser, D., Grable, J., Boyer, H. W., Rosenberg, J. M. & Greene, P. J. (1983). Coordinate ion pair formation between *Eco*RI endonuclease and DNA. *J. Biol. Chem.* **258**, 14638-14646.

Jen-Jacobson, L., Lesser, D. R. & Kurpiewski, M. R., vol. 5 (Springer-Verlag, Berlin, 1991).

Johansen, S., Embley, T. M. & Willasseen, N. P. (1993). A family of nuclear homing endonucleases. *Nucleic Acids Res.* **21**, 4405.

Jordan, R. S. & Pabo, C. O. (1988). Structure of the λ complex at 2.5 Å resolution: details of the repressor–operator interactions. *Science* **242**, 893-899.

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goel, M. & Stevens, T. H. (1990). Protein splicing converts the yeast *TFP1* gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science* **250**, 651-657.

Kim, J., Zwieb, C., Wu, C. & Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* **85**, 15-23.

Kim, Y., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990). Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307-1309.

Kjems, J. & Garrett, R. A. (1991). Ribosomal RNA introns in archae and evidence for RNA conformational changes associated with splicing. *Proc. Natl. Acad. Sci. USA* **88**, 439-443.

Klimasauskas, S., Kumar, S., Roberts, S. J. & Cheng, X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **74**, 357-359.

Kong, H., Roemer, S. E., Waite-Rees, P. A., Benner, J. S., Wilson, G. G. & Nwankwo, D. O. (1994). Characterization of *Bcg*I, a new kind of restriction-modification system. *J. Biol. Chem.* **269**, 683-690.

Koo, H. S. & Crothers, D. M. (1988). Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl. Acad. Sci. USA* **85**, 1763-1767.

Koo, H. S., Drak, J., Rice, J. A. & Crothers, D. M. (1990). Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29**, 4227-4234.

Kostrewa, D. & Winkler, F. K. (1995). Mg^{2+} binding to the active site of *Eco*RV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2Å resolution. *Biochemistry* **34**, 683-696.

Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enymol.* **154**, 367-382.

Kunkel, T. A., Schaaper, R. M., Beckman, R. & Loeb, L. A. (1981). On the fidelity of DNA replication. Effect of the next nucleotide on proofreading. *J. Biol. Chem.* **256**, 9883-98889.

Lambowitz, A. M. (1989). Infectious introns. *Cell* **56**, 232-236.

Lambowitz, A. M. & Belfort, M. (1993). Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**, 587-622.

Lambowitz, A. M. & Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem.* **15**, 440-444.

LeTilly, V. & Royer, C. A. (1993). Fluorescence anisotropy assays implicate protein—protein interactions in regulating *trp* repressor DNA binding. *Biochemistry* **32**, 7753-7758.

Loizos, N., Silva, G. & Belfort, M. (1996). Intron-encoded endonuclease I-*Tev*II binds across the minor groove and induces two distinct conformation changes in its substrate. *J. Mol. Biol.* **255**, 412-424.

Lowery, R., Hung, L., Knoche, K. & Bandziulis, R. (1992). Properties of I-*Ppo*I: A rare-cutting intron-encoded endonuclease. *Promega Notes* **38**, 8-12.

Marshall, P. & Lemieux, C. (1992). The I-*Ceu*I endonuclease recognizes a sequence of 19 basepairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene. *Nucleic Acids Res.* **20**, 6401-6407.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grabble, J. & Rosenberg, J. M. (1986). Structure of the DNA-*Eco*RI restriction complex at 3Å resolution. *Science* **234**, 1526-1541.

McKay, D. B. & Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9Å resolution suggests binding to left-handed B-DNA. *Nature* **290**, 744-749.

Messmore, J. M., Fuchs, D. N. & Raines, R. T. (1995). Ribonuclease A: revealing structure–function relationships with semisynthesis. *J. Am. Chem. Soc.* **117**, 8057-8060.

Michel, F. & Dujon, B. (1986). Genetic exchanges between bacteiophage T4 and filamentous fungi? *Cell* **46**, 323.

Michel, F., Jaquier, A. & Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**, 867-81.

Modrich, P. & Zabel, D. (1976). *Eco*RI endonuclease. Physical and catalytic properties of the homogenous enzyme. *J. Biol. Chem.* **251**, 5866-5874.

Mossing, M. C. & Record, M. T. (1985). Thermodynamic origins of specificity on the *lac* repressor-operator sites. *J. Mol. Biol.* **186**, 295-305.

Mueller, J. E., Bryk, M., Loizos, N. & Belfort, M. (1993). Homing Endonucleases. In *Nucleases* (S. M. Linn, R. S. Lloyd & R. J. Roberts, Ed.), pp. 111-143, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Murdoch, F. E., Grunwald, K. A. & Gorski, J. (1991). Marked effects of salt on estrogen receptor binding to DNA: biologically relevant discrimination between DNA sequences. *Biochemistry* **30**, 10838-10844.

Muscarella, D. E., Ellison, E. L., Ruoff, B. M. & Vogt, V. M. (1990). 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**, 3386-3396.

Muscarella, D. E. & Vogt, V. M. (1989). A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **56**, 443-454.

Nakagawa, K., Morishima, N. & Shibat, T. (1991). A maturase-like subunit of the sequence-specific endonuclease Endo.*Sce*I from yeast mitochondria. *J. Biol. Chem.* **266**, 1977-1984.

Nardone, G. & Chirikjian, J. G. (1987). The enzymes of the *Bam*HI restriction-modification system. *Gene Amplif. Anal.* **5**, 147-184.

Nelson, H. C., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221-226.

Newman, A. K., Rubin, R. A., Kim, S. H. & Modrich, P. (1981). DNA sequences of structural genes for *Eco*RI DNA restriction and modification enzymes. *J. Biol. Chem.* **256**, 2131-2139.

Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1995). Structure of *Bam*HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269**, 656-663.

Nickoloff, J. A., Singer, J. D. & Heffron, F. (1990). *In vivo* analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis. *Mol. Cell. Biol.* **10**, 1174-1179.

Niederweis, M., Lederer, T. & Hillen, W. (1993). An accurate method for determining the helical repeat of DNA is solution reveals differences to the crystal structures of two B-DNA decamers. *J. Mol. Biol.* **228**, 322-326.

Orr-Weaver, T. L., Szostak, J. L. & Rothstein, R. T. (1981). Yeast transformation: a model system for the study of recombination. *Proc. Nat. Acad. Sci. USA* **78**, 6354-6358.

Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. (1983). Genetic applications of yeast transformation with linear and gapped plasmids. *Methods Enzymol.* **101**, 228-245.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmostein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature* **335**, 321-329.

Pabo, C. O. & Lewis, M. (1982). The operator-binding domain of λ repressor: structure and DNA recognition. *Nature* **298**, 443-447.

Perea, J., Desdouets, C., Schapira, M. & Jacq, C. (1993). I-*Sce* III: a novel group I intron-encoded endonuclease from the yeast mitochondria. *Nucleic Acids Res.* **21**, 358.

Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S. & Jannasch,

H. (1992). Intervening sequences in an archaea polymerase gene. *Proc. Natl. Acad. Sci. USA* **89**, 5577-5581.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. (1994). Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22**, 1125-1127.

Perlman, P. S. & Butow, R. A. (1989). Mobile introns and intron-encoded proteins. *Science* **246**, 1106-1109.

Perrin, A., Buckle, M. & Dujon, B. (1993). Asymmetrical recognition and activity of the I-*Sce*I endonuclease on its site and on intron – exon junctions. *EMBO J.* **12**, 2939-2947.

Quirk, S. M., Bell-Pederson, D. & Belfort, M. (1989). Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron reading frames. *Cell* **56**, 455-465.

Radzicka, A. & Wolfenden, R. (1995). A proficient enzyme. *Science* **267**, 90-92.

Record, M. T., Jr., Ha, J. H. & Fisher, M. A. (1991). Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* **208**, 291-343.

Record, M. T., Jr., Lohman, M. L. & De Haseth, P. (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **107**, 145-158.

Resnick, M. A. & Martin, P. (1976). The repair of double strand breaks in the nuclear DNA of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **143**, 119-129.

Riggs, A. D., Suzuki, H. & Bourgeous, S. (1970). The *lac* repressor-operator interaction. I. Equilibrium studies. *J. Mol. Biol.* **48**, 67-83.

Roberts, R. J. & Halford, S. E. (1993). Type II restriction endonucleases. In *Nucleases* (L. S. M., R. S. Lloyd & R. J. Roberts, Ed.), pp. 35-88, Cold Spring Harbor Laboratory Press, New York.

Rocheleau, G. A. & Woodson, S. A. (1994). Requirements for self-splicing of a group I intron from *Physarum polycephalum. Nucleic Acids Res.* **22**, 4315-4320.

Rocheleau, G. A. & Woodson, S. A. (1995). Enhanced self-splicing of *Physarum polycephalum* intron 3 by a second group I intron. *RNA* **1**, 183-193.

Rosenberg, J. M. (1991). Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* **1**, 104-113.

Ruoff, B., Johansen, S. & Vogt, V. M. (1992). Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of *Physarum polycephalum. Nucleic Acids Res.* **20**, 5899-906.

Russell, D. W., Jensen, R., Zoller, M. J., Burke, B., Errede, B., Smith, M. & Herskowitz, I. (1986). Structure of the *Saccharomyces cerevisiae HO* gene and analysis of its upstream regulatory region. *Mol. Cell Biol.* **6**, 4281-4294.

Sargueil, B., Delahodde, A., Hatat, D., Tian, G. L., Lazowska, J. & Jacq, C. (1991). A new specific DNA endonuclease activity in yeast mitochondria. *Mol. Gen. Genet.* **225**, 340-341.

Schapira, M., Desdouets, C., Jacq, C. & Perea, J. (1993). I-*Sce*III an intron-encoded DNA endonuclease from yeast mitochondria. Asymmetrical DNA binding properties and cleavage reaction. *Nucleic Acids Res.* **21**, 3683-3689.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001-1007.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

Selent, U., Rüter, T., Köhler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschläger, T., Wolfes, H., Peters, F. & Pingoud, A. (1992). A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *Eco*RV. *Biochemistry* **31**, 4808-4815.

Shannon, R. D. (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr.* **A32**, 751-767.

Shao, Y., Xu, M.-Q. & Paulus, H. (1995). Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry* **34**, 10844-10850.

Sharma, M., Ellis, R. L. & Hinton, D. M. (1992). 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.* **89**, 6658-6662.

Shub, D., Goodrich-Blair, H. & Eddy, S. (1994). Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.* **19**, 402-404.

Shub, D. A. & Goodrich-Blair, H. (1992). Protein introns: a new home for endonucleases. *Cell* **71**, 183-186.

Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Micheal, F., Tomaschewski, J., Pedersen-Lane, J. & Belfort, M. (1988). Structural conversation among three homologous introns of bacteriophage T4 and group I introns of eukaryotes. *Proc. Natl. Acad. Sci. USA* **85**, 1151-1155.

Siksnys, V. & Pleckaityte, M. (1993). Catalytic and binding properties of restriction endonuclease *Cfr*9I. *Eur. J. Biochem.* **217**, 411-419.

Sinha, N. D., Biernat, J., McManus, J. & Koster, H. (1984). Polymer support oligonucleotide synthesis. XVIII: use of β-cyanoethyl-*N,N*-dialkylamino-/*N*-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA

fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.* **12**, 4539-4557.

Spolar, R. S. & Record, M. T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784.

Starr, D. B. & Hawley, D. K. (1991). The TFIID binds in the minor groove of the TATA box. *Cell* **67**, 1231-1240.

Steitz, T. A. (1990). Structural studies of protein – nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* **23**, 205-280.

Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

Szostak, J. W., Orr-Weaver, T. L. & Rothstein, R. J. (1983). The double-strand-break repair model for recombination. *Cell* **33**, 25-35.

Tanner, N. K. & Cech, T. R. (1985). Self-catalyzed cyclization of the intervening sequence RNA of *Tetrahymena*: inhibition by methidiumprophl•EDTA and localization of the major dye binding sites. *Nucleic Acids Res.* **13**, 7759-7779.

Taylor, J. D., Badcoe, I. G., Clarke, A. R. & Halford, S. E. (1991). *Eco*RV restriction endonuclease binds all DNA sequences with equal affinity. *Biochemistry* **30**, 8743-8753.

Taylor, J. D. & Halford, S. E. (1989). Discrimination between DNA sequences by the *Eco*RV restriction endonuclease. *Biochemistry* **28**, 6198-6207.

Terry, B. J., Jack, W. E., Rubin, R. A. & Modrich, P. (1983). Thermodynamic parameters governing interaction of the *Eco*RI endonuclease with specific and non-specific DNA sequences. *J. Biol. Chem.* **258**, 9820-9825.

Thielking, V., Selent, U., Kohler, E., Wolfes, H., Pieper, U., Geiger, R., Urbanke, C., Winkler, F. K. & Pingoud, A. (1991). Site-directed mutagenesis studies with *Eco*RV

restriction endonuclease to identify regions involved in recognition and catalysis. *Biochemistry* **30**, 6416-6422.

Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. & Schmitz, G. (1991). Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-*Sce*I. *Nucleic Acids Res.* **19**, 189-190.

Thompson, A. J., Yuan, X., Kudlicki, W. & Herrin, D. L. (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.

Thompson, J. F. & Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res.* **16**, 9687-9705.

Thompson, L. D. & Daniels, C. J. (1988). A tRNA(Trp) intron endonuclease for *Halobacterium volcanii*. Unique substrate recognition properties. *J. Biol. Chem.* **263**, 17951-17959.

Thompson, L. D. & Daniels, C. J. (1990). Recognition of exon-intron boundaries by *Halobacterium volcanii* tRNA intron endonuclease. *J. Biol. Chem.* **265**, 18104-18111.

Trautwein, K., Holliger, P., Stackhouse, J. & Benner, S. A. (1991). Site-directed mutagenesis of bovine pancreatic ribonuclease: lysine-41 and aspartate-121. *FEBS Lett.* **281**, 275-277.

Travers, A. A. (1991). DNA bending and kinking-sequence dependence and function. *Curr. Opin. Struct. Biol.* **1**, 114-122.

Turmel, M. J., Boulanger, J., Schnare, M. N., Gray, M. W. & Lemieux, C. (1991). Six group I introns and three internal transcribed spacers in the chloroplast large subuint ribosomal RNA gene of the green algae *Chlamydomonas eugametos*. *J. Mol.Biol.* **217**, 293-311.

Van Cleve, M. D. & Gumport, R. I. (1992). Influence of enzyme–substrate contacts located outside the *Eco*RI recognition site on cleavage of duplex oligodeoxyribonucleotide substrates by *Eco*RI endonuclease. *Biochemistry* **31**, 334-339.

Vipond, I. B., Baldwin, G. S. & Halford, S. E. (1995). Divalent metal ions at the active sites of the *Eco*RV and *Eco*RI restriction endonucleases. *Biochemistry* **34**, 697-704.

Vipond, I. B. & Halford, S. E. (1995). Specific DNA recognition by *Eco*RV restriction endonuclease induced by calcium ions. *Biochemistry* **3466**, 1113-1119.

von Hippel, P. H., Revzin, A., Gross, C. A. & Wang, A., Eds., *Interaction of lac repressor with nonspecific DNA binding sites.* (DeGruyter, Berlin, 1975).

Voorter, C. E. M., de Haard-Hoekman, W. A., van den Oetelaar, P. J. M., H., B. & de Jong, W. W. (1988). Spontaneous peptide bond cleavage in aging alpha-crystallin through a succinimide intermediate. *J. Biol. Chem.* **263**, 19020-19023.

Walsh, C. (1979). *Enzymatic reaction mechanisms*, W.H. Freeman and Company,

Waring, R. B., Davis, R. W., Scazzocchio, C. & Brown, T. A. (1982). Internal structure of a mitochondrial intron of *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. USA* **79**, 6332-6336.

Waugh, D. S. & Sauer, R. T. (1993). Single amino acid substitutions uncouple the DNA binding and strand scission activities of *Fok*I endonuclease. *Proc. Natl. Acad. Sci. USA* **90**, 9596-9600.

Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P. & Butow, R. (1992). Complex recognition site for the group I intron-encoded endonuclease I-*Sce*II. *Mol. Cell. Biol.* **12**, 716-723.

Wernette, C. M., Saldahna, R., Perlman, P. S. & Butow, R. A. (1990). Purification of a site-specific endonuclease, I-*Sce*II, encoded by intron 4α of the mitochondrial *coxI* gene of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 18976-18982.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781-1795.

Withers, B. E. & Dunbar, J. C. (1995). Sequence-specific DNA recognition by the *Sma*I endonuclease. *J. Biol. Chem.* **270**, 6496-6504.

Wittmayer, P. K. & Raines, R. T. (1996). Substrate binding and turnover by the highly specific I-*Ppo*I endonuclease. *Biochemistry* **35**, 1076-1083.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a MAT α2 homeodomain–operator complex suggest a general model for homeodomain–DNA interactions. *Cell* **67**, 517-528.

Woodhead, J. L., Bhave, N. & Malcolm, A. D. (1981). Cation dependence of restriction endonuclease *Eco*RI activity. *Eur. J. Biochem.* **115**, 293-296.

Wu, H. M. & Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-513.

Xu, M.-Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J. & Perler, F. B. (1993). In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell* **75**, 1371-1377.

Xu, M. Q., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y. & Perler, F. (1994). Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J.* **13**, 5517-5522.

Xu, S. Y. & Schildkraut, I. (1991). Cofactor requirements of *Bam*HI mutant endonuclease E77K and its suppressor mutants. *J. Bacteriol.* **173**, 5030-5035.

Xu, S. Y. & Schildkraut, I. (1991). Isolation of *Bam*HI variants with reduced cleavage activities. *J. Biol. Chem.* **266**, 4425-4429.

Zahn, K. & Blattner, F. R. (1987). Direct evidence for DNA bending at the lambda replication origin. *Science* **236**, 416-422.

Zaug, A. J., Grabowski, P. J. & Cech, T. R. (1983). Autocatalytic cyclization of an excised intervening sequence RNA is cleavage-ligation reaction. *Nature* **301**, 578-583.

Zebala, J. A., Choi, J. & Barany, F. (1992). Characterization of steady state, single-turnover, and binding kinetics of the *Taq*I restriction endonuclease. *J. Biol. Chem.* **267**, 8097-8105.

Zhu, H., Macreadie, I. G. & Butow, R. A. (1987). RNA processing and expression of an intron-encoded protein in yeast mitochondria: role of a conserved dodecamer sequence. *Mol. Cell. Biol.* **7**, 2530-2537.