RECEIVED

JAN 30 1996

OSTI

Site Specific Endonucleases for Human Genome Mapping

Final Report

for Period April 1, 1992 - March 31, 1994

Kimberly Knoche, Susanne Selman, Lydia Hung, Jim Hartnett,
and George Golumbeski

Promega Corporation
2800 Woods Hollow Road
Madison, WI 53711

June 1994

Prepared for

The U.S. DEPARTMENT OF ENERGY
Under Award No. DE-FG02-91ER81252.A001

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

ple

pg

RECEIVED

JAN 30 1996

OSTI

ABSTRACT

Current large scale genome mapping methodology suffers from a lack of tools for generating specific DNA fragments in the megabase size range. While technology such as pulsed field gel electrophoresis can resolve DNA fragments greater than 10 megabases in size, current methods for cleaving mammalian DNA using bacterial restriction enzymes are incapable of producing such fragments. Though several multidimensional approaches are underway to overcome this limitation, there currently is no single step procedure to generate specific DNA fragments in the 2-100 megabase size range. In order to overcome these limitations, we proposed to develop a family of site-specific endonucleases capable of generating DNA fragments in the 2-100 megabase size range in a single step. Additionally, we proposed to accomplish this by relaxing the specificity of a very-rare cutting intron-encoded endonuclease, I-Ppo I, and potentially using the process as a model for development of other enzymes.

Our research has uncovered a great deal of information about intron-encoded endonucleases. We have found that I-Ppo I has a remarkable ability to tolerate degeneracy within its recognition sequence, and we have shown that the recognition sequence is larger than 15 base pairs. These findings suggest that a detailed study of the mechanism by which intron-encoded endonucleases recognize their target sequences should provide new insights into DNA-protein interactions; this has led to a continuation of the study of I-Ppo I in Dr. Raines' laboratory and we expect a more detailed understanding of the mechanism of I-Ppo I action to result.

I. Scientific Aims

Current large scale genome mapping methodology suffers from a lack of tools for generating specific DNA fragments in the megabase size range. While technology such as pulsed field gel electrophoresis can resolve DNA fragments greater than 10 megabases in size (1), current methods for cleaving mammalian DNA using bacterial restriction enzymes are incapable of producing such fragments. Though several multidimensional approaches are underway to overcome this limitation, there currently is no single step procedure to generate specific DNA fragments in the 2-100 megabase size range. In order to overcome these limitations, Promega proposed to develop a family of site-specific endonucleases capable of generating DNA fragments in the 2-100 megabase size range in a single step. We proposed to accomplish this by relaxing the specificity of a very-rare cutting intron-encoded endonuclease, I-Ppo I, and potentially using the process as a model for development of other enzymes. In addition, we proposed to examine other intron-encoded endonucleases for potential commercialization.

A. Background

Group I introns are a recently discovered class of mobile genetic elements whose insertion is mediated by an endonuclease encoded within the intron (2). The hallmark of the intron-encoded endonucleases is the large size of their recognition sequences, which range from 15 to 39 base pairs (2). In other respects, the enzymes appear to function very similarly to bacterial type II restriction enzymes, yielding overlapping ends which can be ligated to regenerate the cut site. A number of intron-encoded endonucleases have been cloned and shown to have activity *in vivo* or *in vitro*, including I-Sce I (3) and I-Sce II from yeast (4), I-Tev I and I-Tev II from phage T4 (5), I-Ceu I from *Chlamydomonas eugametos* (6), I-Cre I from *Chlamydomonas reinhardtii* (7), and I-Ppo I from *Physarum polycephalum* (8).

I-Ppo I catalyzes double-strand scissions within a 15 base pair sequence. This sequence (presented in Fig. 1) is based on data from Volker Vogt *et al.* (personal communication and Promega Notes, number 38). The scissions leave 4 base 3'-hydroxyl overhangs and 5'-phosphate recessed ends which can be ligated with T4 DNA ligase (9). Thus, there is no need for any special treatment of DNA digested by I-Ppo I before continuing with standard molecular biological methods.

Calculations demonstrate the potential for advancing current genomic mapping technology by relaxing the specificity of I-Ppo I. (Note: These calculations are simplified for purposes of illustration by assuming an equal distribution of all four bases in human DNA and ignoring the effect of

methylation.) Digestion of total human genomic DNA with a restriction enzyme recognizing an 8 base pair sequence would generate approximately 30,000 fragments with an average size of 100 kb, thus yielding too many fragments to synthesize a large scale map. An enzyme with a 15 base pair recognition sequence, which does not tolerate any base substitutions within this sequence, would theoretically cut the human genome three times. Thus, the cutting frequency of wild type I-Ppo I is too low for genome mapping purposes. However, Table 1 illustrates how reducing the specificity of a "15 base pair cutter" by one base pair increments would alter the number of cut sites per human genome.

Table 1. Comparison of length of recognition sequence with number of cut-sites per human genome.

Length of recognition sequence	# cut sites per human genome
15	3
14	11
13	45
12	178
11	715
10	2,860

Supported by a U.S. Department of Energy Phase I Small Business Innovation Research Grant (#DE-FG02-01ER81252) we were able to overexpress and purify large quantities of I-Ppo I in *E. coli*. Thus, we were able to characterize the physical properties of I-Ppo I. The pH optimum of the I-Ppo I cleavage reaction is 9.5-10.0, and the temperature optimum for this reaction is 37°C. There is an absolute requirement for $MgCl^{++}$. Most importantly, this enzyme appears to be much more robust than the other intron-encoded endonucleases that have been characterized. Using a supercoiled plasmid with a single cut site as a substrate, a specific activity of 8.69×10^6 u/mg was obtained with purified I-Ppo I; this corresponds to a turnover of 2.6 min^{-1} for the dimer (one unit is the amount of enzyme required to cleave 1 μg of substrate to completion in 1 hour). For comparison, the most intensively studied bacterial type II restriction enzyme, Eco RI, has a turnover number of 3.4 min^{-1} (10). In contrast, I-Sce I and I-Sce II (other intron-encoded endonucleases) are much slower, with turnover numbers of 0.058 and 0.011 respectively (11, 12). Therefore, in overall catalytic efficiency, I-Ppo I is very similar to the bacterial type II restriction enzymes, and is superior to other currently available intron-encoded endonucleases. Furthermore, the stability of I-Ppo I at reaction temperature (37°C) is very high, with no detectable loss of activity loss after preincubation of a dilute solution of enzyme (4.5 $\mu\text{g/ml}$) in reaction buffer for 60 minutes. The stability

of I-Ppo I at reaction temperature and the high catalytic efficiency are especially important for genomic mapping due to the prolonged incubation times necessary for diffusion of enzymes into the matrix of an agarose plug.

Another significant advantage of using I-Ppo I for digestion of mammalian DNA was discovered in Phase I: the enzyme is completely insensitive to the presence of methylated cytosines. Most bacterial restriction enzymes will not cleave DNA if there are methylated bases in the recognition sequence (13). Since the DNA methylation pattern varies widely between individuals in any species (13), obtaining reproducible restriction patterns using bacterial restriction enzymes is often problematic. This point is especially important for applications such as the identification of disease-linked alleles in individuals within a population.

Finally, evidence suggests that I-Ppo I cuts DNA with high specificity. In genomes with low complexity (pBR322, lambda, adenovirus 2, PhiX-174, phage T7) there is no evidence of any detectable cleavage. In addition, I-Ppo I does not appear to cut the *E. coli* genome (9). However, I-Ppo I specifically cuts chromosome 12 of the *Saccharomyces cerevisiae* genome within the rDNA repeats, where the I-Ppo I recognition sequence is known to be present. No other cleavage is evident in the yeast genome.

Therefore, Phase I results clearly demonstrated that I-Ppo I possesses the stability, catalytic efficiency, insensitivity to methylation, and performance in an agarose matrix required for commercial success. Moreover, the purification methodology developed during Phase I work was an essential prerequisite for commercialization of this enzyme. Thus, based on all of our initial results, I-Ppo I was chosen as an ideal candidate for further development work intended to relax the enzyme's specificity thereby creating a tool for genomic mapping.

B. Specific Phase II Research Goals and Technical Objectives

The Phase II research goal was to isolate a set of conditions and/or modified I-Ppo I enzymes capable of generating fragments in the 2-100 megabase size range (i.e. 9-13 bp recognition sequence). To accomplish this goal, two general approaches were to be employed. The first approach was to establish extrinsic conditions and identify protein modifications which would lead to decreased specificity of cleavage (presumably due to changes in the conformation of I-Ppo I). The second approach, targeted mutagenesis, was to focus on disrupting specific amino acid-nucleotide base interactions to increase the tolerance of I-Ppo I for base substitutions within its recognition sequence. The success of these methods were to be verified by examining digestion patterns with complex genomic DNA. An additional goal was to

examine other mobile group I intron-encoded endonucleases for potential megabase cutting capabilities and commercialization.

Phase II Technical Objectives

1. Conformational Approach

1.1. Conditions and/or enzyme structural modifications were to be established which lead to relaxed specificity for the I-Ppo I enzyme.

- a. The tolerance for single base substitutions within the 15 base I-Ppo I recognition sequence was to be examined systematically under standard reaction conditions (defined as optimal activity conditions).
- b. External conditions that resulted in relaxed I-Ppo I specificity were to be identified and optimized. These included changes in ionic strength, pH, enzyme concentration, dielectric constant and metal cofactor.
- c. Structural modifications which lead to relaxed I-Ppo I specificity were to be identified. These included covalent modification of buried amino acid residues and limited proteolysis.

Goal 1 was to be considered achieved upon a) determination of tolerance of all single base changes within the I-Ppo I recognition sequence, b) determination of extrinsic conditions and/or structural modifications that cause 1) cleavage at recognition sequences between 9-13 bp., 2) retention of 10% activity of the natural enzyme under standard conditions, 3) $T_{1/2}$ of inactivation greater than 60 minutes at 37° C.

2. Loss of Contact Approach

2.1 The specificity of I-Ppo I was to be broadened by elimination and/or alteration of contacts between amino acids in I-Ppo I and bases in the recognition sequence.

- a. The DNA binding domain of I-Ppo I was to be defined by crosslinking of bases in the recognition sequence to amino acids in I-Ppo I.
- b. Disruption of amino acid-nucleotide contacts were to be achieved by individually replacing each amino acid residue in the DNA binding domain of I-Ppo I with alanine.
- c. The active alanine replacement mutants were to be purified, followed by assessment of specificity and catalytic competency *in vitro*.

Goal 2.1 was to be considered achieved upon a) identification of the amino acid residues in I-Ppo I that contact bases in the recognition sequence, b) isolation of alanine replacement mutants with the following characteristics: 1) cleavage at recognition sequences between 9-13 bp., 2) retention of 10% activity of the natural enzyme under standard conditions, 3) $T_{1/2}$ of inactivation greater than 60 minutes at 37° C.

2.2 Modified I-Ppo I enzymes with relaxed specificity were to be used as targets for additional reductions in specificity.

- a. Covalently modified or mutated I-Ppo I enzymes with relaxed specificity were to be tested under reaction conditions which had been shown to cause relaxed specificity.
- b. Amino acid replacements that resulted in relaxed I-Ppo I specificity were to be combined to generate double mutants.

Goal 2.2 was to be considered achieved upon identification of combinations of conditions and modifications that caused further reductions in specificity.

3. Genomic Mapping Applications

3.1 Conditions and/or modified forms of the I-Ppo I enzyme which yielded relaxed specificity were to be optimized for performance in genomic mapping applications.

- a. Conditions and/or modified I-Ppo I enzymes which yielded relaxed specificity were to be used to digest genomic DNA imbedded in agarose plugs. Pulsed field gel electrophoresis was to be used to assess cutting frequency and overall performance.

Goal 3 was to be considered achieved upon a) demonstration of distinct and reproducible cleavage patterns with genomic DNA using the conditions and modified forms of I-Ppo I described above, b) accurate determination of average fragment size generated by I-Ppo I in forms or under conditions which reduced specificity.

4. Development of Complementary Rare-Cutting Enzymes

Additional mobile group I intron-encoded endonucleases, including I-Cre I (8), were to be examined for their commercial potential. Examination was to include the same studies undertaken for I-Ppo I, including an initial basic assessment of each enzymes' activity and specificity. Pending demonstration of significant potential for megabase cutting capabilities, further efforts were to include detailed analysis of specificity, and, if necessary, efforts to decrease specificity.

Goal 4 was to be considered achieved when those mobile group I intron-encoded endonucleases which demonstrate initial capacity for cutting megabase size fragments of complex genomic DNA, were expressed, purified, and fully analyzed in conditions of actual use through the procedures outlined in Phase I and Phase II of the SBIR research conducted for I-Ppo I.

II. Phase II Research

A. Summary

Promega has commercialized I-Ppo I, and this product has proven useful for genomic analysis, as well as for other applications. Furthermore, current research and development efforts at Promega are exploiting the unique properties of I-Ppo I. For example, the I-Ppo I recognition sequence is being incorporated into a mammalian expression vector, and will serve as a rare-cutting endonuclease site. We expect this vector to be commercially available within one year.

However, in executing the research plan, we encountered many difficulties, all of which demonstrated that the proposed work was much more difficult than anticipated. An increasing body of data, including the work performed during the term of this grant, now shows that I-Ppo I and other intron encoded endonucleases are quite unlike bacterial type II restriction enzymes. Therefore, it is important to remember that nearly all of the research plan was predicated upon assumptions that I-Ppo I studies could be modeled upon previous work with bacterial type II restriction endonucleases.

The most relevant precedent for the conformational approach to relaxing the specificity of I-Ppo I is the induction of star activity in bacterial type II restriction enzymes (14). Star activity is defined as cleavage at sequences which differ from the canonical recognition sequence by a single base pair. Historically, this type of cleavage activity has been induced by: Reaction buffers of suboptimal ionic strength, the presence of organic solvents, high enzyme concentration, and replacement of Mg with Mn (14, 15, 16). In the case of Eco RI, genetic (16), thermodynamic (17), and kinetic (18) data have demonstrated that an allosteric activation step accounts for the discrimination between the canonical sequence and star sites. According to this model, the enzyme binds non-specifically to DNA and slides along the template in an inactive conformation until it locates its cut site. At this specific sequence, the binding energy is sufficient to cause a transition to the active conformation, resulting in cleavage (16). Conditions which promote star activity are believed to shift the conformational equilibrium of Eco RI toward the active conformation, enabling the enzyme to cleave without the full complement of amino acid-base interactions required for tight binding and cleavage at the canonical recognition site.

For the loss of contact approach, the most clearly understood determinant of specificity in protein/DNA interactions is the direct recognition of specific sequences mediated by hydrogen bonding between amino acid side chains and nucleotide bases comprising the DNA molecule (19). It follows from such a digital approach that elimination of amino acid side chains contacting a

given base within a recognition site will eliminate the ability to discriminate between cognate and non-cognate bases at that position (19). This "loss of contact" approach has been used to successfully relax the specificity of other site-specific DNA binding proteins, including CAP (catabolite gene activator protein) (20) and the σ^{70} subunit of *E. coli* RNA polymerase (21).

Despite these elegant precedents, many difficulties were encountered during the research. 1) We were unable to unequivocally determine the tolerance of I-Ppo I for single or multiple base pair substitutions within its recognition sequence. Therefore, we were unable to generate a library of sequences refractory to cleavage by wild type I-Ppo I. Furthermore, the "conformational approach" to relaxing the specificity of I-Ppo I depended on identifying these sequences; thus, we were unable to complete these studies. 2) We found that cleavage of oligonucleotides by I-Ppo I does not match the cleavage of plasmid DNA (and possibly genomic DNA). Thus, I-Ppo I may require positional or spatial cues outside of the central 15 base pair sequence noted in Figure 1. Alternatively, I-Ppo I may have a recognition sequence larger than the 15 base pairs originally proposed. In any case, the discrepancy between the cleavage of oligonucleotide and plasmid sequences created great difficulties in this work. Early findings obtained utilizing the assay became suspect, and progress with this approach became much slower. 3) The "loss of contact" approach depended on crosslinking studies which proved to be very difficult. As a more definitive alternative, crystallography studies have been initiated. However, to date, amino acid-base pair contacts remain unknown. Thus, the "loss of contact" approach has not yet been undertaken.

The many difficulties encountered in this project demonstrate that the intron-encoded endonucleases do not behave like bacterial type II restriction enzymes. Intron-encoded endonucleases are a recently discovered class of proteins; thus, as this project was planned, little was known about these enzymes. Therefore, we were guided by the work done on bacterial type II restriction enzymes, which are by far the best understood group of restriction endonucleases.

As our own work on I-Ppo I progressed, and other workers published their results on intron-encoded endonucleases, it became clear that this new class of endonucleases is quite different from the type II restriction enzymes. Therefore, because of the lack of relevant precedents, it has proven impossible to achieve the commercial goals established for this project. Instead, this work has contributed to furthering the scientific community's understanding of intron-encoded endonucleases.

As it became clear that we would be unable to relax the specificity of I-Ppo I without ample structural data, we decided to concentrate on more tractable goals. Thus, four other intron-encoded endonucleases were examined for

potential commercialization; however, none of these possessed the attractive physical properties of I-Ppo I. Therefore, Promega has decided not to commercialize these proteins. Finally, the ability of I-Ppo I to cleave the human genome was studied to a greater extent than originally planned. From this work, additional uses for this type of enzyme have been discovered.

B. Experimental Results

1. Conformational approach: Establish conditions and/or structural modifications that result in relaxed I-Ppo I specificity through conformational perturbations.

1.1 Systematic examination of the tolerance for single base changes within the recognition sequence.

Before attempting to reduce I-Ppo I specificity, it was necessary to rigorously define the specificity of the wild type enzyme under Promega's standard reaction conditions. When this project was begun, little was known about the tolerance for base pair changes in the recognition sites of intron-encoded endonucleases. Studies on I-Sce I (23) indicated that this enzyme was much like the bacterial type II restriction enzymes in that it did not tolerate single base pair changes. Thus, we used I-Sce I and the type II restriction enzymes as a model for I-Ppo I. Therefore, in order to determine the tolerance of I-Ppo I for base pair changes in its recognition sequence, the ability of I-Ppo I to recognize and cleave all possible single base pair changes within its 15 base pair recognition sequence was undertaken.

a) Assay of single base pair substitutions

The original assay developed to measure the I-Ppo I initial velocity used a 37-base pair double-stranded oligonucleotide as a substrate. This oligonucleotide contained the 15 base pair I-Ppo I recognition sequence at its center, a biotin at one end, and ^{32}P at the other end. After incubation of the oligonucleotide with I-Ppo I, the biotinylated reactants and products were to be complexed by the addition of streptavidin and then the streptavidin-oligonucleotide complexes were to be captured using biotinylated agarose beads. The unbiotinylated, radiolabelled reaction products that were to be released upon I-Ppo I cleavage were to remain in the supernatant and were to be quantitated by liquid scintillation counting.

When the Phase I research ended the assay was not performing adequately. We needed to: 1) reduce the non-specific binding of unbiotinylated reaction products to streptavidin, and 2) further decrease the non-specific binding of the oligonucleotide/streptavidin complex to biotin agarose beads.

During Phase I research, we found that 2.5M NaCl decreased non-specific binding of the oligonucleotide/streptavidin complex to biotin-agarose. However, in subsequent experiments these conditions did not consistently decrease non-specific binding. Additional experiments done to assess different concentrations of NaCl were performed without achieving consistent results. In other attempts to reduce non-specific binding of the oligonucleotides to streptavidin (and to biotin-agarose beads) the following compounds were tested: Triton, Tween-20, acetylated bovine serum albumin, calf thymus DNA, EDTA, and a heat treatment (70°C for 2 min). None of these treatments decreased the non-specific binding sufficiently. During these experiments, capture of the oligonucleotide to streptavidin was not complete or consistent, and capture of the streptavidin/oligonucleotide complex to the biotin agarose beads was also low and inconsistent. Following these results, we abandoned this approach and used the assay described below which separated cleaved from uncleaved double-stranded oligonucleotides by thin layer chromatography.

This assay, which was employed to study I-Ppo I recognition sequences containing one base pair changes (single mutations), relied on homochromatography (23) to separate double stranded oligonucleotide products and reactants. Similar assays have been used successfully for detailed kinetic analyses of bacterial type II restriction enzymes (24). The assay was initially chosen over other *in vitro* or *in vivo* selections because it could rapidly yield quantitative data concerning oligonucleotide cleavage through the use of a phosphorimager.

Forty five pairs of complementary oligonucleotides were synthesized. This "library" of sequences encompassed all three possible base changes at each position of the 15 bp recognition sequence in Fig.1. The oligonucleotides had an extension of five base pairs on both sides of the 15 bp recognition sequence which was shown to be sufficient for efficient cleavage by I-Ppo I. We had initially attempted to use Bam HI recognition sites for these extensions; however, difficulties were encountered in annealing these oligonucleotides because of symmetry. Therefore, one flanking extension contained the five base pair sequence from the Physarum intron homing site, and the other extension contained a random sequence of five base pairs. Oligonucleotides with these extensions did not show any annealing artifacts.

Our intent was to quickly estimate how well the single base pair change sequences were cut by I-Ppo I, followed by careful kinetic studies on the sequences of interest. The following protocol for the assay was used. One single-stranded oligonucleotide of the complementary oligonucleotide pair was 5' end-labeled with ³²P using T4 polynucleotide kinase. The kinase and unincorporated isotope were removed using either a Nensorb column (New

England Nuclear) or a Centriflex primer/probe kit (Advanced Genetic Technologies), and the complementary oligonucleotides were annealed. The double-stranded oligonucleotides were then incubated at 37°C with four different concentrations of I-Ppo I and the buffer control (*i.e.* no enzyme). Thus, each oligonucleotide was used to set up five different reactions. Since the I-Sce I results suggested that I-Ppo I would be unlikely to cut the singly mutant recognition sequences (22), a scheme of ten-fold overdigests was used for the four I-Ppo I treatments. The 1X treatment was based on the conditions needed to cut approximately 90% of the oligonucleotide with the wild type recognition sequence. The 10X, 100X and 1000X treatments were made by using increased enzyme concentrations and/or increased digestion times. Two microliters of each treatment was spotted on a TLC plate and the plates were developed. In each experiment, the wild type oligonucleotide and the one base pair change oligonucleotides were run in parallel. Using a scanning phosphorimager (Molecular Dynamics, 400E), we determined the quantity (pmoles) of oligonucleotide cut and the % of wild type cut (pmoles 1 bp change oligonucleotide cut ÷ pmoles of wild type oligonucleotide cut) for each treatment level used on each 1 bp substitution oligonucleotide. Five different levels of cutting were observed (Table 2).

Table 2. Categories for decreased cutting by I-Ppo I of the 1 bp substitution oligonucleotides.

Categories	% of Wild type Cut	
	< 50%	> 50%
Wild-type		1X, 10X, 100X, 1000X treatments
1-10 fold decrease in cutting	1X treatment	10X, 100X, 1000X treatments
10-100 fold decrease in cutting	1X, 10X treatments	100X, 1000X treatments
100-1000 fold decrease in cutting	1X, 10X, 100X treatments	1000X treatment
>1000 fold decrease in cutting	1X, 10X, 100X, 1000X treatments	

The assay for cutting of the singly mutated oligonucleotides was completed with all possible substitutions for base pairs 5-15 (for numbering of base

pairs see Fig. 1). Most of the 1 bp substitutions were cleaved to a similar degree as the wild type oligonucleotide (Fig. 2). However, a number of oligonucleotides were cut with a 1-10 fold decrease in cutting, and one oligonucleotide was cut with a 10-100 fold decrease in cutting. These results indicate that I-Ppo I tolerates many single base pair changes in its recognition sequence, unlike I-Sce I (22) and the type II restriction enzymes. Since this work was begun, there have been several reports on the specificity of intron-encoded endonucleases. For both I-Sce II (25) and I-Tev I (26), the evidence indicates that these enzymes can tolerate many base pair changes in their recognition sequence. Thus, I-Ppo I is quite like I-Tev I and I-Sce II, and can tolerate much degeneracy in its recognition sequence.

b) Assay with double base pair substitutions

Since I-Ppo I could tolerate most of the 1 bp substitutions in the recognition sequence, we examined several double base pair changes in the oligonucleotide assay to determine the enzyme's ability to cleave doubly substituted sites. The main goal of this analysis was to assess the importance of the I-Ppo I half-site. For Eco RI, if a base pair change is made in one half-site, the enzyme can cut the wild type half-site but not the half-site with the substitution. Therefore, we attempted to assess the importance of the I-Ppo I half-site. We analyzed eight double base pair changes which had either two base pair changes in one half-site or one base pair change in each half-site. In addition, the regions for the base pair changes were either near both ends of the recognition site, or at the ends of the internal hexamer that make up the cleavage site.

All of the double base pair substitution oligonucleotides had decreased cutting when compared to the wild type sequence. As shown in Table 3 decreases in cutting for these oligonucleotides fell into the 1-10 fold, 10-100 fold, 100-1000 fold and >1000 fold categories. These results indicate that I-Ppo I does not tolerate double base pair substitutions as well as most single base pair substitutions.

Table 3. Cutting by I-Ppo I for eight double base pair change oligonucleotides. The double base pair changes in the I-Ppo I recognition site are denoted by bold type.

Decrease in Cutting Category	Double bp Change Recognition Sequence
Wild type	None
1-10 fold	CTCT AG TAAGGTAGC GAGAT C ATTCCATCG ATCTCTTAAGGT AG TAGAGAATTCCAT C T
10-100 fold	ATCTCTTAAGGT ATC TAGAGAATTCCAT AG AGCTCTTAAGGTAGC TC GAGAATTCCATCG CTCT ATT ACGGTAGC GAGAT A AT G CCATCG
100-1000 fold	CTCTCTTAAGGT CTC GAGAGAATTCC AGAG CTCT ATT AA T GTAGC GAGAT A ATT A CATCG
>1,000 fold	ATCTCTTAAGGT CGC TAGAGAATTCC AGCG

Importantly, most doubly mutant oligonucleotides were cleaved to an appreciable extent. This suggests that I-Ppo I does not recognize its target sequence by a "half-site" mechanism. However, given the novelty of this type of analysis applied to intron-encoded endonucleases, these data raised concerns about the validity of the assay. While assays with oligonucleotides worked well for studying the type II bacterial restriction enzymes, oligonucleotide substrates might not be adequate for intron-encoded endonucleases. Thus, we compared cleavage of the oligonucleotides to cleavage of the same sequence within a plasmid and within genomic DNA.

c) Comparison of the I-Ppo I site with base pair changes in an oligonucleotide and a plasmid.

In order to compare cutting of oligonucleotides with cutting of plasmid DNA, the wild type 15 bp recognition sequence and several of the single base pair

substitutions that resulted in decreased cutting were cloned into pGEM-7Zf. Before a comparison was done, the plasmid with the wild type recognition site was tested for the ability to be cut by I-Ppo I. These digests revealed that the plasmid DNA was not cut by I-Ppo I, while p42 (8) (a plasmid containing 42 bp of sequence from the Physarum I-Ppo I intron homing site) and the I-Ppo I "positive control plasmid" currently sold by Promega (a plasmid that also contains the 15 bp recognition sequence) were completely cleaved. Six independent constructs of the plasmid with the wild type recognition site were made. For all of these constructs, cutting by I-Ppo I was severely decreased in comparison to p42 and the "positive control plasmid". However, Afl II (the internal hexamer of the I-Ppo I recognition site is an Afl II site) completely cleaved all 6 constructs, p42, and the "positive control plasmid".

These results suggest that the 15 base pair wild type recognition sequence in the wild type oligonucleotide-plasmid constructs was not adequate for cleavage and/or recognition by I-Ppo I. Because it is only the flanking sequences that differ between the wild type construct, p42, and the "positive control plasmid", our results suggest that the I-Ppo I recognition site may be larger than 15 base pairs (Table 4). Recent results support this notion. DNA footprinting and several protection studies from Volker Vogt's lab (27) indicate that I-Ppo I may be contacting as many as 21 base pairs including six outside the 15 bp recognition sequence. Thus, we currently postulate that the recognition sequence is 16-21 base pairs. Alternatively, it is possible that I-Ppo I requires additional information outside of this region of 16-21 contiguous nucleotides.

Table 4. Flanking sequences for the 15 bp I-Ppo I recognition sequence for three plasmids. p42 contains flanking sequences from Physarum. Base pair changes are denoted in bold type for deviations from the Physarum sequences in the flanking sequences for pGEM/cut-site and the positive control plasmid.

Plasmid	Sequence
pGEM/cut-site	TTG CTCTCTTAAGGTAGC CGGA AAC GAGAGAATTCCATCG GCCT ..
p42	TATGA CTCTCTTAAGGTAGC CAAATG ATACT GAGAGAATTCCATCG GTTTAC
positive control	. TTGG CTCTCTTAAGGTAGC TTAATA . AACC GAGAGAATTCCATCG AATTAT

d) Yeast Mutant Experiments

In recent work Muscarella and Vogt (28) found that induction of I-Ppo I expression in yeast cells (which contain the I-Ppo I recognition sequence in the rDNA) results in nearly complete lethality. However, in each experiment several isolates survived the induction of I-Ppo I expression. The genomic DNA from these strains is not cut *in vitro* by I-Ppo I. Upon careful analysis, these strains were found to have mutations in the I-Ppo I recognition sequence: One mutant had a single base pair change while a second mutant had a double base pair change. By all criteria, these sequences completely abrogate cleavage by I-Ppo I.

To investigate whether altered I-Ppo I recognition sequences are cut similarly in oligonucleotides and in genomic DNA (both *in vivo* and *in vitro*), oligonucleotides with the mutations Muscarella and Vogt (28) described were made and tested for cutting by I-Ppo I in the oligonucleotide assay. The oligonucleotide with the double base pair substitution had a decrease in cutting of 100-1000 fold which appeared consistent with the lack of cutting of genomic DNA (Table 5). In contrast, the oligonucleotide with the single base pair substitution had a decrease in cutting of only 1-10 fold. This small decrease in cutting indicates that the oligonucleotide assay and *in vivo* conditions produce significantly different data. Thus, in order to identify sequences with single and multiple base pair substitutions not cleaved by I-Ppo I we decided to pursue an alternative *in vivo* approach.

Table 5. Cutting by I-Ppo I in the oligonucleotide assay of mutant recognition sequences identified by Muscarella and Vogt (28). Base pair changes in the sequences are denoted by bold type.

Decrease in Cutting Category	Mutant Recognition Sequence
1-10 fold	CTCTCTTAG T GTAGC GAGAGAAT C ACATCG
100-1,000 fold	CTCT A TTAAGGTAGC GAGAT A AATTCCATCG

e) *In vivo* selection procedure to identify sequences that are not cut by I-Ppo I

An *in vivo* selection procedure to identify sequences that are not cut by I-Ppo I was devised using phage M13 containing randomly mutated I-Ppo I recognition sequences and a plasmid that produces I-Ppo I. The constructs required for this screen were made and used as depicted in Fig. 3. This

screen provides a rapid mechanism for direct selection of sequences refractory to I-Ppo I cleavage. In addition, the screen does not rely on cleavage of oligonucleotides and should therefore be unaffected by the problems described above.

M13 phage with randomly mutated I-Ppo I recognition sequences were made by cassette mutagenesis. In this procedure, the mutant I-Ppo I recognition sequences were generated by standard "doped oligonucleotide" synthesis which results in single-stranded oligonucleotides having random base substitutions (either a T, A, C, or G) inserted at various defined positions. The amount of degeneracy used created primarily one, two and three base pair changes within the mutated region. Since our experiments have indicated that I-Ppo I may require more than the 15 bp sequence originally thought to be the recognition sequence, a 41 base pair sequence from *Physarum* that contains the 15 base pair sequence and the flanking DNA from the homing site was cloned into M13. Within this sequence, 40 of the bases were mutated. The mutated single-stranded oligonucleotides were converted to double-stranded DNA using a primer and the Sequenase Version 2.0 T7 DNA polymerase (United States Biochemical). In addition to the mutated recognition site, the oligonucleotide inserted into M13 contained and Eco RI and Bam HI restriction sites at the ends which allowed for directional cloning of the mutated sequences into the Eco RI and Bam HI site of M13mp18. The M13 phage were then used to infect a population of *E. coli* JM109 cells harboring pIPpo. This plasmid is used for production of I-Ppo I and contains the I-Ppo I gene under control of the tac promoter which is inducible by IPTG. When induced, the I-Ppo I should cleave all M13 genomes except those that contain uncleavable I-Ppo I recognition sequences. Thus, only the M13 phage with "resistant" recognition sites should replicate and form plaques. The mutated I-Ppo I recognition site in each of the surviving M13 phage was to then be sequenced to determine the exact site(s) mutated.

Before embarking on the selection procedure with randomly mutated recognition sequences, we began by testing the selection with the M13mp18 containing the wild type I-Ppo I recognition sequence and an M13mp18 without an insert. We made the M13mp18 construct containing the wild type I-Ppo I site by inserting the primer extended oligonucleotide with the wild type sequence. We used various ratios of M13mp18 with the wild type sequence and M13mp18 without the recognition sequence in the selection procedure. However, when *E. coli* JM109 pIPpo cells were grown in the presence of ampicillin (amp), the bacteria failed to produce a healthy lawn. A lawn of *E. coli* JM109 pIPpo was critical for to isolation of M13 plaques.

To examine this result further, *E. coli* JM109 pIPpo were plated on LB-amp plates containing IPTG and LB-amp plates without IPTG (IPTG induces I-Ppo I production). In the presence of IPTG, cell growth was significantly decreased which suggests the overexpression of I-Ppo I may be toxic to *E. coli*.

f) Summary of results for objective 1.1.

For objective 1.1 we wanted to determine how single base substitutions within the I-Ppo I recognition site affect the enzyme's ability to cleave its target. Using the oligonucleotide assay, we examined all possible one base pair substitutions in 11 out of the 15 base pairs. However, identifying the tolerance for base pair substitutions has proven far more difficult than originally anticipated. Clearly, I-Ppo I is more like I-Sce II and I-Tev I in that many degenerate recognition sequences can be tolerated. Cutting of oligonucleotides containing intron-encoded endonuclease recognition sequences does not reflect the ability of these sequences to be cut in genomic DNA *in vitro* or *in vivo*, unlike the type II restriction enzymes. In addition, there is evidence that the I-Ppo I recognition sequence is comprised of 16-21 base pairs rather than the 15 base pairs originally reported. Our results also suggest that overexpression of I-Ppo I may be toxic to *E. coli*. While these results are of significant scientific interest, they made attaining our goals for a commercial product untenable.

Therefore, progress on this project was reviewed at Promega. A rapid *in vitro* assay for the cleavage of mutant oligonucleotides had proven flawed, and rapid *in vivo* screens for "uncleavable" sequences appeared faulty. Moreover, such an *in vivo* screen, in which wild type I-Ppo I would be expressed in *E. coli*, was to be the basis of selecting mutant forms of I-Ppo I. Hence, none of the critical strategies proposed at the outset of this work proved feasible.

We, therefore, concluded that alterations of I-Ppo I activity would have to await detailed structural data. Based on such data, it should be possible to 'rationally' engineer relaxed forms of I-Ppo I; however, during the term of the grant, this was not feasible due to time and financial constraints. Thus, we shifted our focus to purifying large quantities of I-Ppo I to initiate crystallographic studies in the laboratory of Dr. Ronald Raines (University of Wisconsin - Madison). This work is proceeding as described below. Additionally, we began to concentrate on the analysis of additional intron-encoded endonucleases, and on additional applications for I-Ppo I.

1.2. Identification of extrinsic conditions that cause relaxed I-Ppo I specificity.

Since we were unable to identify singly or doubly substituted sequences that are refractory to cleavage by I-Ppo I, we were not able to generate a library of sequences to test extrinsic conditions resulting in altered specificity of I-Ppo I cleavage. Thus, this work was not undertaken.

1.3. Identification of structural modifications contributing to relaxed I-Ppo I specificity.

Same conclusions as stated for section 1.2 immediately above.

2. Loss of contact approach: Relaxation of I-Ppo I specificity through elimination of specific amino acid-base contacts.

2.1 DNA-protein crosslinking studies were to be conducted to define region of protein involved in interaction with recognition sequence.

a) Crosslinking studies

To identify amino acid residues in I-Ppo I that are responsible for its recognition of a specific DNA sequence, cross-linking studies were attempted. In the cross-linking experiments synthetic oligonucleotides containing the photoactive group, 5-bromodeoxyuridine (BrdU) were used. Oligonucleotides were synthesized using the BrdU phosphoramidite to replace thymidines at various positions within the I-Ppo I recognition sequence. Eight oligonucleotides were synthesized: two complementary oligonucleotides without any BrdU-substitutions, six oligonucleotides with single BrdU-substitutions and two complementary oligonucleotides with all eight possible BrdU-substitutions. The oligonucleotides were synthesized on an ABI synthesizer using β -cyanoethyl phosphoramidite chemistry. The oligonucleotides were purified either using C4 reverse-phase column on the HPLC or using OPC columns. The oligonucleotides were end-labeled with ^{32}P , incubated with excess I-Ppo I and irradiated with UV light (254nm). The reactions were then run on a 10% polyacrylamide gel.

Repeated attempts using this approach were unsuccessful in achieving crosslinking. Therefore, we decided to focus on a crystallographic approach. While slower and more labor intensive, this method is capable of providing a maximum of structural information about I-Ppo I. We currently believe that this degree of resolution will be essential for further progress.

b) Crystallography

Crystallographic studies require very large quantities of highly purified protein. Thus, to provide the necessary quantities of I-Ppo I, the purification procedure originally developed was significantly improved. The development of these improvements was noteworthy in that Promega acquired additional "know how" concerning protein purification. Ultimately, 84 mg of highly purified I-Ppo I was prepared and transferred to Dr. Ronald Raines at the University of Wisconsin, and his laboratory is proceeding with the structural studies.

c) Summary of results for objective 2.1

While no data are yet available, the crystallographic studies are now underway.

2.2 Alanine-scanning mutagenesis of DNA binding domain

In an effort to develop mutant I-Ppo I enzymes with relaxed specificity, the amino acid residues that contact the bases in the recognition sequence were to be changed to alanine. While the critical amino acids have not been identified, we subcloned the I-Ppo I gene into a phagemid vector with a tightly repressible promoter; thus, creating a reagent for the site-directed mutagenesis studies. This construct is described below.

a) Subclone the I-Ppo I gene into a phagemid vector with a tightly repressible promoter.

I-Ppo I was subcloned into pSx(T7)T; a vector specifically designed for tight transcriptional regulation of cloned sequences, and efficient incorporation of mismatch mutations. We used the ampicillin sensitive, tetracycline resistant pSx(T7)T plasmid since this vector is designed to facilitate site-directed mutagenesis. pSx(T7)T is a pBR322-derived vector with a cloning region flanked by a phage T7 promoter, a lac operator with two consensus ribosome binding sites on the 5' side, and a T7 transcriptional terminator on the 3' side. Use of this vector in *E. coli* JM109(DE3) provides multiple levels of control of gene expression. Transcription of cloned genes from the T7 promoter occurs only after IPTG induction of the T7 polymerase gene encoded on the lysis-defective lambda lysogen DE3 downstream of a lac UV5 promoter. In addition, negative control is exerted by binding of lac repressor protein to the lac operator region between the T7 promoter of pSx(T7)T and the cloned gene. The level of lac repressor protein normally present in *E. coli* can be titrated out by multicopy plasmids carrying lac operator regions resulting in inefficient repression of downstream genes. The pSx(T7)T

expression system overcomes this problem in two ways. First, JM109 contains a *lac* I^q allele that overproduces the *lac* repressor approximately ten-fold. Second, because pSx(T7)T is derived from pBR322, it exists in a relatively low copy number (about 20 per cell). The combination of positive and negative regulatory elements in the pSx(T7)T system allows repression levels of greater than 1000-fold using luciferase as a marker, and typically very low basal levels of expression are observed. Another advantage of the pSx(T7)T vector is that upon induction, relatively high levels of expression are attained so that mutant I-Ppo I genes isolated in the selection can be expressed and the protein purified for further analysis without subcloning. (Please see the schematic of pSx(T7)T/I-Ppo presented in Fig. 4.)

The I-Ppo I gene was cloned into pSx(T7)T using standard methods. A DNA fragment with the I-Ppo I gene was generated using the polymerase chain reaction (PCR) and following restriction digestion, the fragment was directionally cloned into the Bam HI and Xho I sites of pSx(T7)T. These plasmids were transformed into *E. coli* JM109 and the constructs were verified by diagnostic restriction enzyme digests. The constructs containing the I-Ppo I gene then were transformed into *E. coli* JM109(DE3).

Once this construct was made, independent clones were sequenced to confirm that the correct sequences were present for the promoter, ribosome binding sites, translational start sites and I-Ppo I structural gene. Several constructs were found to have the correct sequences for all needed elements. However, we discovered that the actual I-Ppo I gene was longer by 25 amino acids than the previously reported sequence. The I-Ppo I gene in pIPpo (the vector from which the PCR product was generated) (see Fig. 5) was also sequenced to confirm that the longer version of the I-Ppo I gene was correct. The identical sequence for I-Ppo I was found in both pIPpo and the pSx(T7)T/I-Ppo. This finding had great significance for our own work, and will be important for others working on I-Ppo I. See Fig. 6 for revised sequence of I-Ppo I gene.

b) Summary of Objective 2.2

Since we have not identified amino acid residues that interact with the bases in the I-Ppo I recognition sequence, we were unable to attempt alanine-scanning mutagenesis of the DNA binding domain. However, we did construct pSx(T7)T/I Ppo I and found that the actual I-Ppo I gene was longer than previously reported. Thus, data and reagents have been produced to expedite this work as soon as the structural information becomes available.

2.3 Alanine replacement mutations that result in relaxed I-Ppo I specificity were to be combined to further relax specificity.

This work was not undertaken.

2.4. Combined conformational and loss of contact approach: Further reduction in specificity through combined approaches.

This work was not undertaken.

3. Genomic Mapping Applications: Digestion of genomic DNA and analysis by pulsed field gel electrophoresis.

3.1. I-Ppo I and mutant enzymes were to be tested for human mapping applications by analyzing the digested products of various complex DNA templates.

Since no mutant I-Ppo I enzymes or conditions that result in relaxed specificity were obtained, we attempted to further characterize the usefulness of wild type I-Ppo I as a tool for complex genome analysis. Our studies were intended to identify useful applications for very rare cutting restriction enzymes.

a) Cutting of the human genome by I-Ppo I

I-Ppo I cleaves ribosomal DNA in eukaryotic genomes, specifically in the ribosomal DNA (rDNA) repeats. In the Phase I grant, we found that the *S. cerevisiae* genome is cleaved only in the rDNA repeats on chromosome XII. In the Phase II work, we analyzed human genomic DNA to determine whether I-Ppo I cleaves human rDNA repeats. To accomplish this, DNA from several human cell lines was digested with restriction enzymes, resolved by pulse field gel electrophoresis (PFGE) followed by ethidium bromide staining, and by hybridization with an oligonucleotide probe (17-mer) containing the I-Ppo I recognition sequence. To increase the probability of detection, a human erythroleukemic cell line containing 60 chromosomes (on average) per nucleus, was employed in this work.

The rationale in these experiments was to verify the presence of the I-Ppo I recognition site in the human genome by hybridization of a probe containing the I-Ppo I recognition sequence to human DNA digested with other restriction enzymes (Not I and Csp I). As a confirmation, the same restricted DNA was secondarily cleaved by I-Ppo I, separated by gel electrophoresis, then probed with an oligonucleotide containing the I-Ppo I recognition sequence. Hybridization conditions were used such that if I-Ppo I sites were cleaved in the second digest, the signal against the probe would no longer be detectable.

Unexpectedly, for DNA digested with I-Ppo I, a 35-39 kb digestion product was visible by ethidium staining after PFGE. This result suggests that there

are multiple I-Ppo I sites in the human genome. From the Southern blot results, DNA digested with Not I or Csp I, and probed with the 17 base I-Ppo I recognition sequence reveals multiple bands. This results confirms the existence of I-Ppo sites flanked by Not I (or Csp I) sites. When Not I digested or Csp I digested K-562 DNAs are secondarily digested with I-Ppo I, no signals are detectable after hybridization with the same probe. This result indicates that the DNA fragments which hybridize to the 17 base probe contain an I-Ppo I recognition site.

The observation that human DNA is cut by I-Ppo I was extended by analyzing two additional human cell-lines DNAs (Raji and Colo320). When genomic DNA from these cell lines and K-562 was digested with Hae III, resolved on gels, and probed with the 17 base sequence, a prominent 0.2 kb hybridization signal was observed. However, when the DNAs were secondarily digested with I-Ppo I, no hybridization was detected indicating that the 0.2 kb fragment contains an I-Ppo I site.

b) Localization of I-Ppo I sites in the human genome

GenBank databases were searched for homologies between the I-Ppo I recognition sequence and the published human genome sequences. Two sequences which matched perfectly were found: 1) human 28S ribosomal DNA, a single-copy gene of 5,025 bp (29); and 2) the human 28S orphon rDNA fragment, an 823 bp sequence (30) corresponding to nucleotides 3627-4105 of the human 28S rDNA gene.

To investigate the chromosomal localization of the I-Ppo I recognition sites in the human genome, human-rodent hybrid cell lines containing individual human chromosomes were analyzed by hybridization. Two probes were designed for these experiments. Probe A, a 300 bp probe from the 28S orphon region which included the I-Ppo I site was used to locate all I-Ppo I sites within specific human chromosomes. Probe B, a 290 bp probe from the 28S rDNA repeats which excludes the region known to be in the orphon was used to locate all rDNA loci. Clones for production of these probes were provided by Dr. Iris Gonzales.

DNA from all cell lines were completely digested with Pvu II, resolved on 0.9% agarose gels, transferred to membranes and hybridized with Probe A or Probe B. Cell lines examined in hybridization experiments were Coriell panel #2 (contains the entire set of human chromosomes); Coriell 10660, Coriell 10479, and Bio SM909 (all containing chromosome 14); and Coriell 10478 (contains chromosome 20).

In hybridizations using Coriell panel #2, Probe A (used to locate I-Ppo I sites) hybridized to chromosomes 13, 15, 20, 21 and 22 indicating that the I-Ppo I

site was located in these chromosomes. Comparing our results with those of Henderson *et al.* (31), two differences were observed: 1) chromosome 14 did not hybridize but contains rDNA repeats and 2) chromosome 20 did hybridize but does not contain rDNA loci.

To investigate the absence of hybridization to chromosome 14 by Probe A, three cell lines containing chromosome 14 (Bio SM909, Coriell 10660, and Coriell 11535) were hybridized to Probe A and B separately. We found that two cell lines (Bio SM909 and Coriell 11535) gave positive signals with both probes. Cell line 10660 was inconclusive which may be due to low copy number of chromosome 14. This confirms the expected localization of I-Ppo I sites on chromosome 14 where rDNA loci have been described by Henderson *et al.* (31). The negative result from the hybridization of probe A with the cell line containing chromosome 14 in the Coriell panel #2 may indicate that there is a sequence deletion in rDNA loci in this hybrid cell line.

To investigate the hybridization of Probe A to chromosome 20, cell line Coriell 10478 which contains chromosome 20 was tested with both probes separately. Probe A hybridized to this cell line while Probe B did not. This suggests that chromosome 20 may harbor the 28S orphion rDNA fragment and that the I-Ppo I site may be distributed in locations apart from the rDNA loci.

To further study this possibility, *in situ* hybridizations were attempted in collaboration with Dr. Hsiao-chen Chang (at the University of California - San Diego); however, these experiments proved inconclusive.

c) Summary of objective 3.1

I-Ppo I proved to be a useful tool in analyzing the distribution of rDNA loci within the human genome. The presence of a rare sequence cleaved by a restriction enzyme proved to be diagnostic for a particular gene. Moreover, taken together, our findings suggest that a DNA sequence containing the I-Ppo I site may be mobile within the human genome. Different cell lines appear to be heterogeneous for the presence of the orphion sequence within chromosome 14. While this finding could also be ascribed to polymorphisms or aberrations within cultured cells, the corresponding finding that chromosome 20 harbors a 28S orphion sequence provides more compelling evidence for the mobility of this sequence. Thus, studies with I-Ppo I allowed us to rapidly study the distribution of specific sequences within the human genome. The strategy we employed may have utility for other investigations of this kind.

4. Development of other intron-encoded endonucleases for genomic mapping purposes.

Because the experiments to alter the properties of I-Ppo I were proceeding very slowly and with much difficulty, we shifted the emphasis of our work and began to study the commercial potential of other intron-encoded endonucleases. However, by this time, Promega had been selling I-Ppo I for over one year. Our sales figures were disappointing, and clearly suggested that wild type intron encoded endonucleases appeal to a very limited market segment.

a) I-Cre and I-Dir

Both I-Cre I, an intron-encoded endonuclease from *Chlamydomonas reinhardtii* and I-Dir I, and intron-encoded endonuclease from *Didymium iridis*, were examined in detail. *E. coli* clones expressing these two proteins were obtained from academic scientists. Efforts to purify and characterize each enzyme revealed that neither I-Cre I nor I-Dir I possess the attractive physical properties of I-Ppo I. Moreover, we were unable to purify these proteins with the same degree of ease encountered in purifying I-Ppo I. These results, combined with our sales figures for I-Ppo I, led to Promega's decision not to commercialize these enzymes.

b) Intron-encoded endonucleases from thermophilic Archaea

Two other intron-encoded endonucleases from thermophilic archaea, were evaluated under terms of a secrecy agreement for use in a novel application. These enzymes will be referred to as enzyme A and enzyme B. These enzymes have optimal activity at 65°C, making them potentially useful in the novel assay. Thus, the enzymes were tested for their ability to cut double-stranded DNA and to make single-stranded nicks at their recognition sites under non-optimal conditions.

The preparation of enzyme A was a heat treated lysate produced from an *E. coli* clone expressing enzyme A. Unfortunately, this preparation contained a contaminating exonuclease, active at 65°C, which digested double-stranded but not single-stranded DNA. Moreover, the enzyme A preparation not only cut the plasmid containing its recognition site but also five other plasmids that did not contain the recognition site. The "contaminating" endonuclease activity could be: 1) an activity from the *E. coli* lysate that has a small recognition site, or 2) enzyme A may tolerate a very great deal of substitution in its recognition sequence making the enzyme cut much more frequently. However, based on our experience, few *E. coli* enzymes survive heating to 65°C; thus, we decided not to commercialize enzyme A.

The enzyme B preparation had been purified to homogeneity, and this enzyme cut only the double-stranded plasmid that contained its recognition sequence. However, this enzyme was not able to make single-stranded nicks in DNA. Thus, the enzyme was not useful in the novel assay under development. Furthermore, while active at 65°C, this enzyme is virtually inactive at 37°C-42°C. Thus, it has very little utility in standard molecular cloning techniques.

c) Summary of Objective 4

Four intron-encoded endonucleases were tested for feasibility for commercialization, but none showed the characteristics necessary for the desired degree of commercial success.

5. Conclusions

Our research has uncovered a great deal of information about intron-encoded endonucleases. We have found that I-Ppo I has a remarkable ability to tolerate degeneracy within its recognition sequence, and we have shown that the recognition sequence is larger than 15 base pairs. These findings are extremely promising from a technical standpoint, because they suggest that a detailed study of the mechanism by which intron-encoded endonucleases recognize their target sequences should provide new insights into DNA-protein interactions. However, the difficulties encountered indicate that the commercial goals will be extremely challenging, or even out of reach in the foreseeable future.

We are very excited about the basic scientific questions raised by this research. The study of I-Ppo I is continuing in Dr. Raines' laboratory and we expect a more detailed understanding of the mechanism of I-Ppo I action. However, current market projections for the products that we intended to develop from I-Ppo I indicate that the size of the market does not justify the long-term human resource and financial commitment required. Thus, further effort on I-Ppo I appears to be an excellent scientific endeavor, but of lower than expected commercialization potential due to marketing/technological shifts during the period of this research. Therefore, we do not intend to actively pursue continued development of enhanced I-Ppo I activities. Instead, we have agreed to support Dr. Raines' research effort on the questions raised by this study and to exploit any commercializable technologies resulting from his work. Because of the fundamental importance of this research, we intend to publish our results in the near future. Promega greatly appreciates the support of the Department of Energy which allowed us to undertake this long-term, basic research project.

III. Literature Cited

1. J.D. Levine & C.L. Cech (1989) *Biotechnology* 7;1033
2. P.S. Perlman & R.A. Butow (1989) *Science* 246;1106
3. A. Gauthier, M. Turmel & C. Lemieux (1991) *Curr. Genet.* 19;43
4. A. Delahodde, V. Goguel, A.M. Becam, F. Creusot, J. Perea, J. Banroques & C. Jacq (1989) *Cell* 56;431
5. D. Bell-Pettersen, S. Quirk, J. Clyman & M. Belfort (1990) *Nucl. Acids Res.* 18;3763
6. M. Turmel, J. Boulanger, M.N. Schnare, M.W. Gray & C. Lemieux (1991) *J. Mol. Biol.*
7. F. Durrenberger & J.D. Rochaix (1991) *EMBO J.* 10;3495
8. D. Muscarella & V. Vogt (1989) *Cell* 56;443
9. D.E. Muscarella, E.L. Ellison, B.M. Ruoff & V. Vogt (1990) *Molec. Cell. Biol.* 10;3386
10. P. Modrich & D. Zabel (1976) *J. Biol. Chem.* 251;5866
11. C. Monteilhet, A. Perrin, A. Thierry, L. Colleaux & B. Dujon (1990) *Nucl. Acids Res.* 18;1407
12. C.M. Wernette, R. Saldanha, P.S. Perlman & R.A. Butow (1990) *J. Biol. Chem.* 265;18976
13. M. McClelland & M. Nelson (1988) *Gene* 74;291
14. J.M. Rosenberg (1991) *Curr. Opinion in Struct. Biol.* 1;104
15. F. Barany (1988) *Gene* 65;149
16. J. Heitman & P. Model (1990) *EMBO J.* 10;3369
17. J.-H. Ha, R.S. Spolar & M.T. Record (1989) *J. Mol. Biol.* 209;801
18. B.J. Terry, W.E. Jack & P. Modrich (1985) *J. Biol. Chem.* 13130
19. R. Ebright (1991) *Meth. Enzymol.* 208;620
20. R. Ebright, A. Kolb, H. Buc, T. Kunkel, J. Krakow & J. Beckwith (1987) *Proc. Natl. Acad. Sci.* 84;6083
21. T. Gardella, H. Moyle & M. Susskind (1989) *J. Mol. Biol.* 206;579
22. L. Colleaux, L.D. D'Auriol, F. Galibert & B. Dujon (1988) *Proc. Natl. Acad. Sci.* 85;6022
23. C.R. Aiken & R.I. Gumport (1991) *Meth. Enzymol.* 208;433
24. V. Thielking, J. Alves, A. Fliess, G. Maass & A. Pingoud (1990) *Biochemistry* 29;4682
25. C. Wernette, R. Saldanha, D. Smith, D. Ming, P.S. Perlman & R.A. Butow (1992) *Molec. Cell. Biol.* 12;716
26. M. Bryk, S.M. Quirk, J.E. Mueller, N. Loizos, C. Lawrence & M. Belfort (1993) *EMBO J.* 12;2141
27. E.L. Ellison & V. Vogt (1993) *Molec. Cell. Biol.* 13;7531
28. D. Muscarella & Vogt (1993) *Molec. Cell. Biol.* 13;1023
29. I.L. Gonzalez, J.L. Gorski, T.J. Campen, D.J. Dorney, J.M. Erickson, J.E. Sylvester & R.D. Schmickel (1985) *Proc. Natl. Acad. Sci.* 82;7666
30. J. Munro, R.H. Burdon & D.P. Leader (1986) *Gene* 48;65

31. A.S. Henderson, D. Wasburton & K.C. Atwood (1972) Proc. Natl. Acad. Sci. 69;3394

Figure 1. Fifteen base pair I-Ppo I recognition sequence; slashes indicate cut sites.

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'
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		

Figure 2. Decrease in cutting of the one base pair change oligonucleotide at positions 5-15. For decrease in cutting categories see Table 2. Base changes correspond to the sequence in the top strand. The cells marked X correspond to the wild type base.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
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'
base change																
A					1- 10X	W.T.	W.T.	X	X	W.T.	W.T.	W.T.- 10X	X	W.T.	W.T.	
T		X		X	W.T.	X	X	W.T.	W.T.	W.T.- 10X	W.T.	X	1- 100X	W.T.	W.T.	
C	X		X		X	1- 10X	W.T.- 10X	1- 10X	W.T.	W.T.	W.T.	1- 10X	10- 100X	1- 10X	X	
G					W.T.	1- 10X	1- 10X		W.T.	X	X	W.T.	1- 10X	X	W.T.

Figure 3. Schematic diagram of the genetic selection for I-Ppo I sites not cleaved by wild type I-Ppo I.

Expression of I-Ppo I from the pIPpo vector is dependent upon addition of IPTG to the culture. Upon induction, I-Ppo I should cleave all M13 genomes except those which contain I-Ppo I recognition sequences not recognized and/or cleaved by wild type I-Ppo I. Therefore, only M13 phage with "resistant" recognition sites should replicate and form plaques. This provides a powerful selection that should be easy to score. The essential features of each vector are shown in the Figure and discussed in the text.

Figure 4. Features of pSx(T7)T/IPpo.

This plasmid was constructed by directional cloning of the PCR product generated by amplification of the I-Ppo I gene. The plasmid confers tetracycline resistance and ampicillin sensitivity. As discussed in the text this construction should allow for extremely tight repression of I-Ppo I expression in the absence of IPTG.

Figure 5. Features of pIPpo.

This plasmid confers ampicillin resistance and tetracycline resistance. Upon addition of IPTG, I-Ppo I expression is driven by the tac promoter (P_{tac}).

A

350

tgg gag tca cta gac gac aac aaa ggc aga aac tgg tgc ccg ggt cca acg ggg gat gtg tcc atg cgg tgg ttt gtt taa
W E S L D D N K G R N W C P G P T G D V S M R W F V OCH

B

350

tgg gag tca cta gac gac aac aaa ggc aga aac tgg tgc ccg ggt Ccc aac ggg gga tgt gtc cat gcg gtg gtt tgt
W E S L D D N K G R N W C P G P N G C V H A V V C

tta agg cag ggt ccg ttg tac ggc ccg ggc ccg ggc ggc act gtg gca ggt cct caa caa agg ggc agt cac ttt gtg gta taa
L R Q G P L Y G P G A T V A G P Q Q R G S H F V D OCH

Figure 6. A: The nucleotide and deduced amino acid sequence for the I-Ppo I gene as published by Muscarella *et al.* (Mol. Cell. Biol. 10 :3386-3396). B: The nucleotide and deduced amino acid sequence obtained at Promega during work under this grant. The bold, italicized, upper case C (C) was missing from the published sequence. Addition of this base create a +1 frameshift mutation which removes the original stop codon (in bold, underlined, lower case: taa). This alteration results in a predicted protein that is 25 amino acids longer than originally predicted. The "new" stop codon is shown in bold, lower case letters (**taa**). The sequence shown above commences with base #350 in the numbering scheme used by Muscarella *et al.* (see p. 3389 of the citation given above).