ELSEVIER

Gene 206 (1998) 11–21

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Degenerate DNA recognition by I-*Ppo*I endonuclease

Paula K. Wittmayer, Jennifer L. McKenzie, Ronald T. Raines *

Department of Biochemistry, University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706-1569, USA

Received 31 March 1997; received in revised form 5 September 1997; accepted 14 September 1997; Received by C.M. Kane

Abstract

The I-*Ppo*I endonuclease is encoded by a group I intron found in the slime mold *Physarum polycephalum*. To initiate homing of its encoding intron, I-*Ppo*I catalyzes a specific double-stranded break within a 15-bp recognition site. The high substrate specificities of I-*Ppo*I and other homing endonucleases make these enzymes valuable tools for genomic mapping and sequencing. Here, we report on the ability of I-*Ppo*I to cleave recognition sites that contain a wide variety of mutations generated randomly or deliberately. We find that much degeneracy is tolerated within the recognition site of I-*Ppo*I. Few single substitutions prevent cleavage completely. In addition, many sites with multiple substitutions are cleaved efficiently. In contrast, deletions or insertions within the I-*Ppo*I recognition site are detrimental to catalysis, indicating that proper registry between the protein and its substrate is critical. Finally, we find that the sequence of the flanking regions can influence catalysis by I-*Ppo*I. Thus, I-*Ppo*I has both the complex binding specificity of a transcription factor and the catalytic ability of a restriction endonuclease. © 1997 Elsevier Science B.V.

Keywords: DNA cleavage; Genome mapping; Intron homing; Restriction endonuclease; Substrate specificity; Transcription factor

1. Introduction

Specificity has evolved between enzymes and substrates. This specificity ensures that enzymes catalyze only particular reactions in metabolic pathways. The evolution of substrate specificity has been especially exquisite in enzymes that operate on nucleic acids. For example, *Escherichia coli* DNA polymerase III holoenzyme and other polymerases with 3′→5′ exonuclease proofreading activity replicate double-stranded DNA with only one mistake per 10^8 bp (Kornberg and Baker, 1992). The *Eco*RV restriction endonuclease cleaves its recognition site at a rate greater than 10^6-fold faster than non-cognate sites, and the alteration of a single base within the recognition site is sufficient to cripple the enzyme's catalytic activity (Taylor and Halford, 1989). Illumination of the chemical devices used by enzymes to achieve such extraordinary levels of molecular discrimination is an area of intense investigation.

The non-covalent interactions used to achieve the high substrate specificity of homing endonucleases are not understood. 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). Homing is a gene conversion event that results in the transfer of a copy of the coding sequence of the endonuclease to a cognate intronless allele. Homing endonucleases cleave double-stranded DNA within large (15–40 bp) asymmetrical recognition sites and tolerate varying degrees of degeneracy within the recognition sites (Mueller et al., 1993). Much interest has been focused on the specificity of these homing endonucleases, as they have a potential value for genomic mapping and sequencing projects.

The I-*Ppo*I endonuclease is a 2×20-kDa homodimer that mediates homing of the third intron found in the gene encoding 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-bp homing site (Fig. 1) (Ellison and Vogt, 1993). Cleavage within this sequence generates four-base 3′-hydroxyl overhangs. In the absence of Mg^{2+}, the enzyme binds specifically to its substrate and bends it 38° from linearity, but does not catalyze cleavage (Wittmayer and Raines, 1996). In addition, the enzyme makes contact with regions flanking the 15-bp recognition site (Ellison and Vogt, 1993). The footprint of I-*Ppo*I, found by using the enzymic probe deoxyribonuclease and the small-molecule probe dimethyl sulfate (DMS), extends to 23 bp.

* Corresponding author. Tel: +1 608 262 8588; Fax: +1 608 262 3453; e-mail: raines@biochem.wisc.edu

PII S0378-1119(97)00563-5

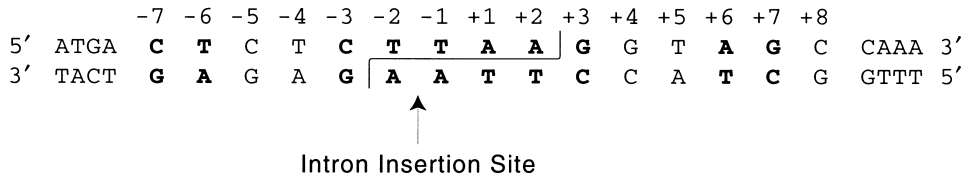

Fig. 1. Homing site of the I-*Ppo*I endonuclease. The DNA bases are numbered about the dyad of pseudosymmetry. The bases that contribute to the symmetry are in bold. The central 15 bp comprise the recognition site as determined by deletion analysis (Ellison and Vogt, 1993). The bases outside the central 15 bp are the flanking regions protected in the I-*Ppo*I footprint. The arrow denotes the location of Pp LSU3 intron insertion, and the line indicates the cleavage site. All 23 bp represent potential sites of substitution in our studies.

In this study, we examine in detail the specificity of the I-*Ppo*I endonuclease. In vitro, we tested the catalytic activity of I-*Ppo*I with DNA recognition sites containing mutations that span the entire 23 bp footprint. From this information, we find 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. Many sequences with two substitutions are cleaved with equivalent or only slightly reduced efficiency from that of the wild-type sequence. Also, few single point substitutions thwart catalysis. In contrast to substitutions, single deletions or insertions within the I-*Ppo*I recognition site are detrimental to catalysis, indicating that proper spacing of contacts along the protein–DNA interface is critical. Finally, we find that bases flanking this 15-bp recognition site can affect the activity of I-*Ppo*I.

2. Materials and methods

2.1. Bacterial strains, plasmids, enzymes, and chemicals

Escherichia coli strain XL1-Blue {*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac* [(F′ *proAB*, *lacI*^{q}*ZDM15*, Tn *10* (tet^{r})]} and pBluescript(SK+) were from Stratagene (La Jolla, CA). Plasmid p42 was a generous gift from Promega (Madison, WI) (Muscarella et al., 1990). All enzymes for the manipulation of DNA were from Promega. 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 (St. Louis, MO). Agarose was from Life Technologies (Gaithersburg, MD). Bacto yeast extract, Bacto tryptone, and Bacto agar were 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 an *E. coli* expression system, was a generous 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) with 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 (1.0 mM), NaCl (0.20 M), and glycerol (10% v/v) and were diluted into solutions of bovine serum albumin (0.2 mg/ml) and glycerol (10% v/v), prior to individual assays.

2.2. General methods

The oligonucleotides in Table 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 the manufacturer. All sequences were determined by dideoxynucleotide chain-termination sequencing using the Sequenase Version 2.0 Quick Denature kit and the M13 forward primer from United States Biochemical (Cleveland, OH). Recombinant DNA techniques were performed as described (Ausubel et al., 1989).

2.3. Random mutagenesis of the *I*-Ppo*I* recognition site

Substrates for I-*Ppo*I contained single, double, or multiple substitutions within the I-*Ppo*I recognition site. The oligonucleotide PW48 was synthesized using a degenerate mixture of 95% of the wild-type nucleotide and 1.7% of each of the remaining three nucleotides at each position in the underlined region Table 1. The non-underlined sequence was synthesized with 100% of the wild-type nucleotide. Using the oligonucleotide PW47 as a primer, the second strand of the recognition site was synthesized by treatment with dNTPs and the Klenow fragment of DNA polymerase I. The resulting double-stranded oligonucleotides were cleaved with the *Sal*I and *Xba*I endonucleases, and the fragments were ligated into pBluescript(SK+) that had also been cleaved with *Sal*I and *Xba*I. The ligation mixtures were transformed by electroporation into *E. coli* strain XL1-Blue. The total number of recombinants obtained (8×10^4) was estimated by plating 10% of the liquid

Table 1
Synthetic oligonucleotides used to create mutant substrates for I-*Ppo*I

Oligo	Sequence
PW47	CGCTCTAGAGTAACT
PW48[a]	CGCTCTAGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTGTCGACACCG
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]Underlined region was synthesized with the degenerate mixture described in the text.
[b]H represents an equal mixture of A, C, and T. B represents an equal mixture of G, T, and C.

culture on to LB agar containing ampicillin (0.10 mg/ml), IPTG (0.5 mM), and X-GAL (0.10 mg/ml). Individual colonies were picked from this library, and liquid cultures (1.5 ml) of these colonies were grown in LB medium containing ampicillin (0.10 mg/ml). Plasmid DNA was isolated from these cultures, sequenced, and screened as a substrate for I-*Ppo*I.

2.4. 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 reflects the 5′ end of the recognition site and PW52/PW53 reflects the 3′ end of the recognition site. These oligonucleotides were phosphorylated by treatment with ATP and T4 polynucleotide kinase. Complementary oligonucleotides were annealed by heating a solution (1 μM of PW50 and PW51, or PW52 and PW53) of 10 mM Tris–HCl buffer, pH 8.0, containing NaCl (0.20 M) to 90°C, and then cooling the solution to room temperature. The annealed oligonucleotides were ligated into pBluescript(SK+) that had been cleaved with *Sal*I and *Xba*I. Sequences were verified by dideoxynucleotide sequencing.

2.5. Synthesis of supplementary recognition sites

Synthetic oligonucleotides PW55–PW59 were created to increase the number of available singly substituted recognition sites. 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 1) and two rounds of the PCR performed on a template that contains the wild-type sequence (the plasmid p42, which contains the recognition site of I-*Ppo*I). The first round of the 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 the primers in the second round of the PCR, again with p42 as the template. The resulting mutagenic, double-stranded oligonucleotide was purified, cleaved with *Sal*I and *Xba*I, and ligated into pBluescript(SK+) that had been cleaved with *Sal*I and *Xba*I. Sequences were verified by dideoxynucleotide sequencing.

2.6. *I*-Ppo*I endonuclease assays*

To assess the mutant recognition sites as substrates for I-*Ppo*I, individual plasmids (0.25 μg) containing the sites were incubated with I-*Ppo*I (0.25 ng, 2.5 ng, or 25 ng) in 15 μl of 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM) and DTT (1.0 mM) for 1 h at 37°C. Under these reaction conditions, the catalytic activity of I-*Ppo*I is maximal (Lowery et al., 1992), and 100% of the wild-type site was cleaved at all three I-*Ppo*I concentrations. Reactions were quenched by the addition of a solution (5 μl) of SDS (1% w/v), EDTA (0.05 M), glycerol (10% v/v), and Bromophenol Blue (0.01% w/v). Products were separated by electrophoresis in a TAE agarose (0.75% w/v) gel.

To distinguish the mutant sites as substrates for I-*Ppo*I, they were classified into four types. Class 1 sites were unaffected; these substrates were cleaved by I-*Ppo*I as thoroughly as was the wild-type site. Class 2 sites were affected slightly; 70–90% of these substrates were cleaved at the lowest enzyme concentration but 100% were cleaved at the other two enzyme concentrations. Class 3 sites were affected moderately; 100% of these substrates were cleaved only at the highest I-*Ppo*I con-

centration. Class 4 sites were affected greatly; these substrates were not cleaved completely at even the highest enzyme concentration (Fig. 2).

3. Results

We have examined the influence of mutations to its recognition site on catalysis by the I-*Ppo*I endonuclease. Base pairs located within the 15-bp 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 substitutions. 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 substitutions (14%), 20 were double substitutions (26%), 19 were triple substitutions (25%), 10 were quadruple substitutions (13%), two were quintuple substitutions (3%), and five were insertions or deletions (6%). This distribution of substitutions was not as anticipated. The predicted distribution of substitutions from a particular doping frequency can be determined from Eq. (1) (Hermes et al., 1990):

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

where m is the length of the doped region, a 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 substitutions, 21% double substitutions and 8% triple substitutions. Our population was under-represented by sequences with single substitutions and over-represented by those with multiple substitutions, suggesting that the actual doping frequency was close to 10%.

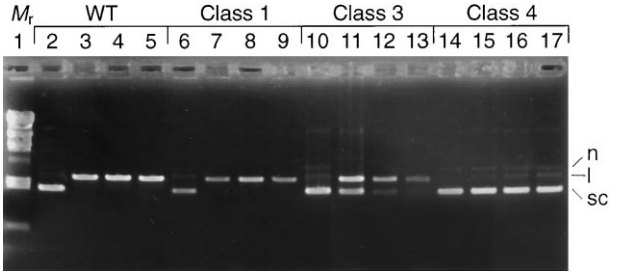

Fig. 2. Activity assay of mutant substrates with the I-*Ppo*I endonuclease. Plasmid substrates were incubated with increasing concentrations of enzyme in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM) and DTT (1.0 mM) for 1 h at 37°C. Lane 1 has λ DNA digested with *Hin*dIII, lanes 2–5 have the wild-type substrate; lanes 6–9 have a class 1 substrate; lanes 10–13 have a class 3 substrate; and lanes 14–17 have a class 4 substrate. Lanes 2, 6, 10, and 14 have no endonuclease; lanes 3, 7, 11, and 15 have 1.0 nM endonuclease; lanes 4, 8, 12, and 16 have 0.010 μM endonuclease; and lanes 5, 9, 13, and 17 have 0.10 μM endonuclease. Different forms of the plasmid are indicated as n, nicked; l, linear; and sc, supercoiled.

To enlarge the population of sites containing single substitutions, additional clones were constructed individually. These supplementary sequences are denoted with an S- prefix. Of the final population of 102 clones screened, 10 were wild type (10%), 23 were single substitutions (23%), 30 were double substitutions (30%), 19 were triple substitutions (19%), 10 were quadruple substitutions (10%), two were quintuple substitutions (2%), and eight were insertions or deletions (8%).

The 102 plasmids of the final population were tested as substrates for I-*Ppo*I. Gels from typical assays are shown in Fig. 2. Sites were grouped into one of four categories, depending on whether catalysis by I-*Ppo*I was unaffected (Class 1), affected slightly (Class 2), affected moderately (Class 3), or affected greatly (Class 4). Clones that were constructed independently and that contain the same recognition sites displayed similar activities, demonstrating the validity of the assay and classification.

3.1. Singly substituted sites

The effect of single substitutions on the activity of I-*Ppo*I is shown in Table 2. Nineteen independent single substitutions within the I-*Ppo*I footprint region were assayed. The substitutions outside the 15-bp recognition site (sequences 5, 67, and 72) have no effect on catalysis by I-*Ppo*I. These sequences with flanking region substitutions are cut as efficiently as is the wild-type sequence. In addition, three single substitutions (sequences 60, S15, and S17) have no effect on catalysis (Class 1). These substitutions, A(+1) to G, T(−4) to G, and C(−5) to A, are within the 15-bp recognition site (Table 2). Only two of the single substitutions (sequences 74 and S10) prevent catalysis by I-*Ppo*I (Class 4). These sequences have a C(−3) to A or A(+6) to C substitution (Table 2). Cleavage of the remaining sites with single substitutions is affected only slightly or moderately (Class 2 or Class 3, respectively). These results indicate that degeneracy is tolerated within the 15-bp recognition site of I-*Ppo*I. A summary of the results for the sites with a single substitution is shown as a matrix in Fig. 3.

3.2. Deletions and insertions

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 sites. The effect of these mutations on catalysis by I-*Ppo*I is shown in Table 3. Two isolated clones contained insertions (sequences 6 and 23); six clones contained deletions (sequences 32,

Table 2
Effect of single base-pair substitutions on catalysis by I-*Ppo*I

Identification	Sequence (5′→3′)[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[c]	ATGA CTCT**A**TTAAGGTAGC CAAA	4
S23 (and S24)	ATGA CTCT**T**TTAAGGTAGC CAAA	2
S17	ATGA CTC**G**CTTAAGGTAGC CAAA	1.5[d]
S14	ATGA CT**T**TCTTAAGGTAGC CAAA	2
S15	ATGA CT**A**TCTTAAGGTAGC CAAA	1
S11 (and S18)	ATGA C**A**CTCTTAAGGTAGC CAAA	2.5[d]
S19	ATGA C**C**CTCTTAAGGTAGC CAAA	3
S4	ATGA **A**TCTCTTAAGGTAGC CAAA	2
S5	ATGA **T**TCTCTTAAGGTAGC CAAA	2.5[d]
S6 (and 75)	ATGA **A**TCTCTTAAGGTAGC CAAA	2

[a]Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are based on cleavage assays with plasmid substrates incubated for 1 h at 37°C in 25 mM CAPS/CHES buffer, pH 10, containing $MgCl_2$ (10 mM), DTT (1 mM), and I-*Ppo*I endonuclease (1.0 nM, 0.010 μM, or 0.10 μM). Class 1, unaffected; Class 2, affected slightly; Class 3, affected moderately; Class 4, affected greatly.
[c]This sequence was also isolated by Muscarella and Vogt (1993), who showed that it was not cleaved by I-*Ppo*I in vivo.
[d]When the results of individual assays varied, the average result is given.

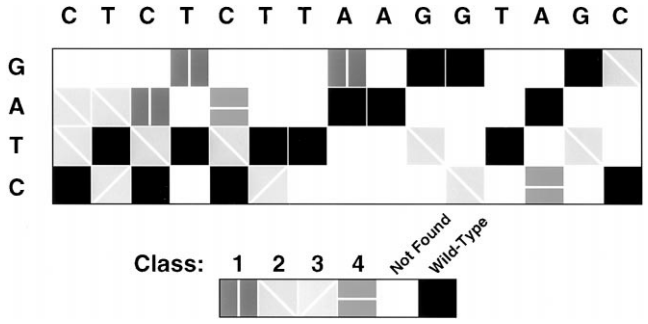

Fig. 3. Summary of sites examined with single base substitutions. The data in Table 2 are summarized as a matrix. The wild-type recognition sequence is given (top). The three possible base substitutions are shown (left). Black boxes represent the wild-type sequence. White boxes indicate base substitutions for which no singly substituted sequence was found. Shaded regions represent the results for the base substitutions for which singly substituted sequences were found.

40, 44, S12, S20, and S22). All of these sequences were poor substrates for the enzyme (Class 4). These results suggest that proper registry of protein–DNA contacts is critical for catalysis by I-*Ppo*I.

3.3. Flanking region substitutions

The effect of substitutions in the flanking regions on catalysis is readily apparent from the data shown in Tables 4 and 5. As stated above, single substitutions within the flanking regions do not affect catalysis by I-*Ppo*I—such substrates are cleaved as well as is the wild-type sequence. Multiple substitutions within only the flanking regions, however, have varied effects on catalysis by I-*Ppo*I (Table 4). Five of the seven sequences that contain multiple substitutions outside the undisturbed 15-bp recognition site display slightly (Class 2, sequences 15, 54 and 71) or moderately (Class 3, sequences 4 and 39) reduced cleavage. In addition, the effect of a single substitution within the 15-bp recognition site could be enhanced when combined with a substitution in the flanking regions (Table 5). Again, five of seven sites with single substitutions (sequences 14, 41, 43, 49 and 58) are worse substrates when the flanking region too is altered. Interestingly, the effect of flanking regions is not absolute. In two of the seven sequences (sequences 42 and 48), substitutions in the flanking regions do not alter the effect of the single substitutions within the recognition site (Table 5). These results reveal that substitutions outside the 15-bp recognition site can modulate catalysis by I-*Ppo*I.

3.4. Symmetric sites

The recognition site of I-*Ppo*I contains partial symmetry (Fig. 1). Ten of the 15 bp within this recognition site contribute to an imperfect palindrome. Many restriction

Table 3
Effect of base-pair insertions and deletions on catalysis by I-*Ppo*I

Identification	Sequence (5′→3′)[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
[c]	ATGA CTCTCT(**T**)TAAGGTAGC CAAA	
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 shown in bold in parentheses. Deleted bases are represented by 'Δ'. Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.
[c]This sequence was isolated by Muscarella and Vogt (1993), who showed that it was not cleaved by I-*Ppo*I in vivo.

Table 4
Effect of base-pair substitutions in the flanking regions on catalysis by I-*Ppo*I

Identification	Sequence (5′→3′)[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 CA**G**A	2
54	AT**C**A CTCTCTTAAGGTAGC **T**AAA	2
71	A**G**GA CTCTCTTAAGGTAGC **T**AA**T**	2
4	**G**TGA CTCTCTTAAGGTAGC **A**AA**G**	3
39	**T**TG**C** CTCTCTTAAGGTAGC CA**G**A	3
Single base-pair substitutions		
5	ATGA CTCTCTTAAGGTAGC C**G**AA	1
67	ATG**T** CTCTCTTAAGGTAGC CAAA	1
72	ATGA CTCTCTTAAGGTAGC CA**G**A	1

[a]Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.

Table 5
Effect of single base-pair substitutions combined with subsitutions in the flanking regions on catalysis by I-*Ppo*I

Identification	Sequence (5′→3′)[a]	Activity class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
Effect by flanking sequence substitutions		
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 **A**AG**A**	4
57	ATGA CTCTCTTAAGCTAGC CAAA	2
No effect by flanking sequence substitutions		
48	A**A**GA CT**A**TCTTAAGGTAGC CAA**C**	1
S15	ATGA CT**A**TCTTAAGGTAGC CAAA	1
42	ATG**G** CTCTCTTAAGGTAGG CA**T**A	2
12	ATGA CTCTCTTAAGGTAG**G** CAAA	2

[a]Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.
[c]When the results of individual assays varied, the average result is given.

endonucleases form dimers and recognize perfect palindromes, with each monomer of the enzyme recognizing one half-site (Roberts and Halford, 1993). Because the I-*Ppo*I endonuclease also functions as a dimer (Ellison and Vogt, 1993), we examined the catalysis by I-*Ppo*I on the symmetric sites composed of the two wild-type half-sites. Both of these symmetric sites are cleaved as well as is the wild-type sequence (Class 1). This result is consistent with a dimeric I-*Ppo*I endonuclease recognizing the partial symmetry of the wild-type recognition site.

3.5. Multiply substituted sites

Three types of effects are observed with the multiply substituted substrates: tolerance, cumulative effects, and non-cumulative effects (Tables 6–8, respectively). First, the tolerance of I-*Ppo*I for degeneracy within its recognition site is readily apparent with multiply substituted substrates. Table 6 shows that numerous sequences with multiple substitutions 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 many interactions that are lost because of substitutions within the substrate. The protein appears to be able to maintain or maximize favorable interactions and reduce or minimize unfavorable interactions.

Second, the effect of multiple substitutions within the same substrate can be cumulative. A comparison of the single and multiple substitutions in which the effect on the activity of I-*Ppo*I from multiple substitutions is larger than that from any of the single substitutions is shown in Table 7. In this data subset, I-*Ppo*I exhibits a graded discrimination between wild-type and mutant sequences. Single substitutions alone have no effect, or only a slight effect, on cleavage, but in combination with an additional substitution, cleavage is slowed to a greater extent. For example, the double mutant that contains an A(+1) to T substitution and a C(−3) to T substitution (sequence S25) is a Class 3 mutant, but the single C(−3) to T substitution (sequence S23) is a Class 2 mutant, and two sequences with an A(+1) to T substitution (sequences 46 and 62) are Class 1 mutants. The sequences in Table 7 are examples of substitutions that are cumulative; together, the substitutions exert a greater effect on catalysis by I-*Ppo*I than they do alone.

The third small set of multiple substitutions is shown in Table 8. The effect of these substitutions is non-cumulative; the substitutions combined on the same substrate exert an effect no greater than the most debilitating substitution. For example, the triple mutant that contains a T(−6) to C substitution, an A(+1) to G substitution, and an A(+10) to C substitution (sequence 76), is a Class 3 mutant. This sequence is cleaved no more slowly than the single T(−6) to C substitution (sequence S19). Despite the presence of substitutions additional to the T(−6) to C, the cleavage of the substrate is not slowed to a great extent.

Finally, although the activity of I-*Ppo*I is maximal at pH 10, the enzyme is also active near netural pH (Lowery et al., 1992; Wittmayer and Raines, 1996). Under extreme solution conditions, restriction endonucleases are capable of cleaving sequences that are similar, but not identical, to their defined recognition sequence. This altered specificity is termed 'star' activity (Polisky et al., 1975). To determine whether the degenerate specificity of I-*Ppo*I is due to I-*Ppo*I* activity, we assayed the cleavage of three sequences (31, 45, and 46; Table 6) in the buffer used in our previous study

Table 6
Effect of multiple base-pair substitutions on catalysis by I-*Ppo*I: substitutions that have no or a minimal effect

Identification	Sequence (5′→3′)[a]	Activity class[b]
WT	ATGA CTCTCTTAAGGTAGC CAAA	1
46	ATGA C**G**CTCTT**T**AGGTAGC CAAA	1
48	A**A**GA CT**A**TCTTAAGGTAGC CAA**C**	1
42	ATG**G** CTCTCTTAAGGTAG**G** CA**T**A	2
2	**G**TG**T** CTCTCTT**C**AGGTAGC C**TT**A	1
35	A**G**GA CT**T**TCT**A**AAGGTAGC CAAA	2
9	ATGA CTCTCTTAAGG**CAT**C CA**T**A	2
31	ATGA CTCGCTTAAGG**C**AG**T** CAAA	2
45	ATGA CT**A**T**TT**AAAGGTAGC CAA**G**	2
33	ATGA CTCTC**G**TAAGGTAGC CA**T**A	1
62	ATGA CTCTCTT**T**AGGTAGC CA**C**A	1
41	ATGA CTCTCTT**G**AGGTAGC CAA**C**	2
66	ATGA CTC**A**CTTAAGGTAGC CAA**G**	2

[a]Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.

Table 7
Effect of multiple base-pair substitutions on catalysis by I-*Ppo*I: substitutions that have a cumulative effect

Identification	Sequence (5′→3′)[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 (and S24)	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 (and 75)	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	ATGA 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 shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.
[c]When the results of individual assays varied, the average result is given.

Table 8
Effect of multiple base-pair substitutions on catalysis by I-*Ppo*I: substitutions that have a non-cumulative effect

Identification	Sequence (5′→3′)[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**AT**C** CA**T**A	2
8	ATGA CTCTCTTAAGGTA**T**C CAAA	2

[a]Substituted bases are shown in bold. The four 5′ and 3′ bases are in the flanking regions.
[b]Activity classes are as described in Table 2.

(Wittmayer and Raines, 1996), which was 10 mM Tris–HCl buffer, pH 7.5, containing NaCl (50 mM), $MgCl_2$ (10 mM), DTT (1 mM), and bovine serum albumin (50 μg/ml). At pH 10, the multiple base-pair substitutions in sequences 31, 45, and 46 have no, or a minimal, effect (Table 6). No increase in specificity was observed at the more neutral pH (data not shown). Apparently, the degenerate specificity of I-*Ppo*I is not due to 'star' activity.

3.6. Composite analysis of entire data set

Together, the data in Tables 2–8 show that sequence recognition by I-*Ppo*I is complex. Many single and multiple substitutions are tolerated within the recognition site, but others are not. To summarize the impact of substitutions within the recognition site of I-*Ppo*I, we used the data collected for all mutant substrates to create the matrix shown in Fig. 4. In this matrix, each base substitution at each position is considered individually. All sequences containing a particular substitution are considered together as a subset. The base substitution is assigned a classification based on the substrate that is cleaved most readily. For example, if an individual base substitution is a component of three different mutant substrates of Class 2, 3, or 4, the individual base substitution is assigned a classification of 2. If an individual base substitution is a component of two different substrates of either Class 3 or Class 4, the individual substitution is assigned a classification of 3. The classification scheme in Fig. 4 is less accurate than that in Fig. 3 for substitutions that appear only in multiply substituted sequences. Thus, the matrix in Fig. 4 is only qualitative, reporting on the highest activity observed for a substrate with a particular substitution.

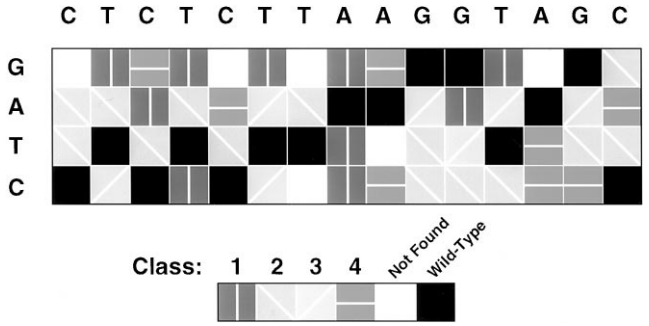

Fig. 4. Summary of all sites examined. The data in the supplementary material are summarized as a matrix, as in Fig. 3. The classification of individual bases from data with multiple substitutions provide an estimate of a maximal effect. For bases that are class 1 or 2, this estimate is likely to be reasonable. For bases that are class 3 or 4, this estimate is less likely to be reasonable unless the substitution was analyzed as a singly substituted sequence (Fig. 3). If the class 3 or 4 base is represented in red was analyzed only in a substrate with multiple substitutions, the loss of catalysis by I-*Ppo*I could be due to other substitutions within the sequence or to a cumulative effect (Table 7).

4. Discussion

This study is the first extensive analysis of the substrate specificity of the I-*Ppo*I endonuclease. We have found that in vitro, I-*Ppo*I has a more relaxed specificity than was thought previously (Lowery et al., 1992). Sequences with single, double, and multiple base-pair substitutions within the homing site of I-*Ppo*I are acceptable substrates for the enzyme (Tables 2 and 6). The tolerance for degeneracy by I-*Ppo*I within its recognition site 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 site (Marshall and Lemieux, 1992; Wernette et al., 1992; Bryk et al., 1993; Durrenberger and Rochaix, 1993).

The level of degeneracy accepted by the different homing endonucleases spans a 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 base-pair substitutions and affected moderately by an additional 11 single base-pair substitutions within its 18-bp recognition site (Colleaux et al., 1988). I-*Ceu*I cleavage activity is eliminated by 11 single substitutions within its 19-bp recognition site (Marshall and Lemieux, 1992). The homing endonucleases, I-*Sce*II and I-*Cre*I, are less stringent. Of the 36 single base-pair substitutions examined within the 18-bp recognition site of I-*Sce*II, 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 that of I-*Sce*II (Durrenberger and Rochaix, 1993). Single and multiple mutations are tolerated by this homing endonuclease. The T4 phage endonuclease, I-*Tev*I, is more permissive. No single base pair within the 48 bp spanning the homing site of I-*Tev*I is essential for cleavage (Bryk et al., 1993). Our data suggest that I-*Ppo*I has intermediate specificity, falling between I-*Sce*I and I-*Tev*I. I-*Ppo*I also appears to be slightly more specific than I-*Sce*II.

Two single substitutions of the 16 examined substitutions within the 15-bp recognition site of I-*Ppo*I, C(−3) to A (sequence S10) and A(+6) to C (sequence 74), were shown to prevent the cleavage activity by I-*Ppo*I (Class 4; Table 2). These two positions (+6 and −3) are presumably key regions for recognition by I-*Ppo*I. Position +6 is partially protected in the minor groove during footprinting experiments with DMS, and the neighboring G(+7) is strongly protected in the major groove (Ellison and Vogt, 1993). This information in conjunction with the cleavage data on the sequence with a single substitution at position +6 (sequence 74), suggests that this base pair is an important protein–DNA contact. Interestingly, I-*Ppo*I does not appear to contact the −3 position closely as G(−3), the complementary base of C(−3), is not protected during footprinting experiments with DMS (Ellison and 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) base pair with an AT base pair 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 an indirect readout for its recognition.

The flanking regions outside the 15-bp recognition site (Fig. 1) can modulate cleavage by I-*Ppo*I. Some sequences with multiple substitutions in only the flanking regions were poor substrates for I-*Ppo*I (Table 4). Likewise, cleavage of some sequences with single internal substitutions were hindered to a greater extent when combined with substitutions in the flanking regions (Table 5). Flanking sequences are known to have similar modulating effects in other DNA-binding proteins. For example, cleavage by the HO nuclease is affected by regions outside the core sequence (Nickoloff et al., 1990). Most point mutations outside this core region of approximately 8 bp have little effect on activity, whereas multiple changes outside this core region severely diminish cleavage by the homing 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 influence the rate of cleavage for the proximal half-site by the *Eco*RI restriction endonuclease (Van Cleve and Gumport, 1992). The footprint of the

I-*Ppo*I endonuclease shows that flanking regions are protected in the I-*Ppo*I–substrate complex (Ellison and Vogt, 1993). From our data, the interactions at these positions appear to influence the cleavage efficiency by the I-*Ppo*I endonuclease.

Few single substitutions were observed to prevent completely catalysis by I-*Ppo*I. A majority of the individual single substitutions within the 15-bp recognition site only had a slight effect on cleavage (Table 2). These results suggest that individual base-pair interactions make only small contributions to specificity. Maintaining interactions with each DNA base pair within the recognition site apparently is not necessary for catalysis by I-*Ppo*I. Remaining protein–DNA interactions can compensate for any lost contacts between the enzyme and the 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 and Record, 1985).

The effect of multiple substitutions on catalysis by I-*Ppo*I is varied. Multiple substitutions can have a cumulative or non-cumulative effect. For example, single substitutions 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 7). This cumulative effect of multiple substitutions would suggest a more global mode of sequence recognition. Still, other single substitution combinations can result in no alteration of the effect on cleavage (Table 8). These combinations of substitutions have no greater effect than that of the most debilitating single substitution alone.

Together, these results portray a clear difference between I-*Ppo*I endonuclease and the type II restriction endonucleases. I-*Ppo*I does not maintain the strict specificities of most restriction endonucleases. Instead, I-*Ppo*I appears to recognize its substrates in a manner more reminiscent of transcription factors, which tend to maintain sequence preferences rather than strict sequence requirements (Mossing and Record, 1985). I-*Ppo*I is thus a distinct protein with the complex binding specificity of transcription factors superimposed on the catalytic ability of restriction endonucleases.

Acknowledgement

P.K.W. is supported by Biotechnology Training Grant GMO8349 (NIH). J.L.M. is a Mary Shine Peterson undergraduate scholar. We are grateful to Promega Corp. for their generous support of this work, and to S.J. Mannino for helpful discussions.

References

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

Bryk, M., Quirk, S.M., 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.

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.

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

Durrenberger, 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.

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.

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.

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

Kornberg, A., Baker, T.A., 1992. DNA Replication, 2nd ed. W.H. Freeman, New York.

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 base pairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene. Nucleic Acids Res. 20, 6401–6407.

Mossing, M.C., Record, M.T., Jr, 1985. Thermodynamic origins of specificity in the lac repressor–operator interaction. Adaptability in the recognition of mutant operator sites. J. Mol. Biol. 186, 295–305.

Mueller, J.E., Bryk, M., Loizos, N., Belfort, M., 1993. Homing endonucleases. In: Linn, S.M., Lloyd, R.S., Roberts, R.J. (Eds.), Nucleases. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 111–143.

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., 1993. A mobile group I intron from *Physarum polycephalum* can insert itself and induce point mutations in the nuclear ribosomal DNA of *Saccharomyces cerevisiae*. Mol. Cell Biol. 13, 1023–1033.

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.

Polisky, B., Greene, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M., Boyer, H.W., 1975. Specificity of substrate recognition by the *Eco*RI restriction endonuclease. Proc. Natl. Acad. Sci. USA 72, 3310–3314.

Roberts, R.J., Halford, S.E., 1993. Type II restriction endonucleases. In: Linn, S.M., Lloyd R.S., Roberts, R.J. (Eds.), Nucleases. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 35–88.

Sinha, N.D., Biernat, J., McManus, J., Koster, H., 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. 1984. Nucleic Acids Res. 12, 4539–4557.

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

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.

Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S., Butow, R.A., 1992. Complex recognition site for the group I intron-encoded endonuclease I-*Sce*II. Mol. Cell Biol. 12, 716–723, 1903.

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