© 1998 Oxford University Press
Nucleic Acids Research, 1998, Vol. 26, No. 19 **4409–4412**

Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1

Motomu Nishioka, Shinsuke Fujiwara, Masahiro Takagi and Tadayuki Imanaka[1,*]

Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan and [1]Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto 606-8501, Japan

Received June 22, 1998; Revised and Accepted August 11, 1998

DDBJ/EMBL/GenBank accession no. D29671

ABSTRACT

Two intein endonucleases, denoted PI-*Pko*I and PI-*Pko*II, in the DNA polymerase gene of the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 were expressed in *Escherichia coli* and the recombinant endonucleases were characterized. Both endonucleases were thermostable and cleaved their intein-less DNA sequences leaving four base 3′-hydroxyl overhangs. PI-*Pko*I exhibited 22 times higher specific activity than PI-*Pko*II and the activity of PI-*Pko*II was enhanced at higher potassium ion concentrations (1 M). Recognition sequences were also determined using synthetic oligonucleotides inserted into plasmid pUC19. It was shown that DNA sequences of 19 and 16 bp are needed for cleavage by PI-*Pko*I and PI-*Pko*II, respectively. PI-*Pko*II could cleave the downstream junction region between intein-encoding and mature DNA polymerase regions and cleavage by PI-*Pko*II could be detected even when chromosomal DNA of *P.kodakaraensis* KOD1 was used as substrate. Therefore, it is suggested that these endonucleases are switching endonucleases whose function lies in the rearrangement of chromosomal DNA.

INTRODUCTION

Protein splicing is a post-translational reaction involving precise excision of an intervening protein sequence, termed an intein, from a precursor protein and subsequent ligation of the external protein segments to form a native peptide bond (1–5). Two of the intriguing properties of this reaction are that protein splicing involves autocatalytic excision of the intein and that an excised intein often exhibits site-specific endonuclease activity which recognizes and cleaves the intein-less DNA allele. Since the first discovery of protein splicing in the *TFP1* gene (also designated *VMA1*) encoding the 69 kDa catalytic subunit of the vacuolar H^+-ATPase in *Saccharomyces cerevisiae* (6,7), protein splicing has been reported from all three phylogenetic domains: bacteria, eukarya and archaea (8–12). In previous studies, four of the known inteins have been shown to possess endonuclease activity and among them the *Sce VMA* intein endonuclease from *S.cerevisiae*, named PI-*Sce*I, has been particularly well studied (13–16). The PI-*Sce*I endonuclease exhibits 34% amino acid identity to the *S.cerevisiae* HO endonuclease, an enzyme that mediates the switching of mating type in yeast, with the homology being greatest in the conserved dodecapeptide sequences corresponding to the active sites of these endonucleases (6). This dodecapeptide sequence, called the LAGLIDADG motif, is shared not only by intein and HO endonuclease but is also found in homing endonucleases encoded by group I and archaeal introns (17–20). Enzymes possessing the LAGLIDADG motif cleave DNA within their recognition sequences to leave four base 3′-hydroxyl overhangs. The recognition sequences are generally asymmetrical and long, with sizes of 12–40 bp (17).

We have shown that the thermostable DNA polymerase gene from a hyperthermophilic archeon *Pyrococcus kodakaraensis* KOD1 contains two intervening sequences (21). In the present study, we describe the characterization of these two new thermostable endonucleases, PI-*Pko*I and PI-*Pko*II.

MATERIALS AND METHODS

Plasmids used

Two DNA fragments coding for KOD DNA polymerase, each containing one of two intein sequences, were constructed using PCR techniques as explained in a previous report (21). Each of the amplified fragments was inserted into an expression vector (pET 8c) and the resultant plasmids, pET-*pol*(*intein-1*) for PI-*Pko*I and pET-*pol*(*intein-2*) for PI-*Pko*II, were used to transform *Escherichia coli* BL21(DE3).

Expression and purification of recombinant proteins

Gene expression of *E.coli* cells harboring pET-*pol*(*intein-1*) or pET-*pol*(*intein2*) was induced by addition of 1 mM IPTG at mid-exponential phase and the cells were harvested after 4 h incubation by centrifugation (8000 *g* for 10 min). The cell pellet was resuspended in buffer A (10 mM Na-phosphate, pH 7.0, 0.1 M NaCl, 0.1 mM EDTA, 1 mM DTT) and disrupted by sonication,

*To whom correspondence should be addressed. Tel: +81 75 753 5568; Fax: +81 75 753 4703; Email: imanaka@sbchem.kyoto-u.ac.jp

Figure 1. SDS–PAGE of purified recombinant PI-*Pko*I and PI-*Pko*II.

Figure 2. N-Terminal sequence of purified PI-*Pko*I and PI-*Pko*II. Conserved amino acid residues observed at protein splicing junctions are indicated by an asterisk. The amino acid residues determined experimentally in this study are underlined.

Figure 3. Endonuclease activity measurement for PI-*Pko*I and PI-*Pko*II. Lane 1, DNA size marker (sizes shown in bp on the right); lane 2, pUCSS04 digested with *Sca*I and PI-*Pko*I; lane 3, pUCSS04 digested with *Sca*I and PI-*Pko*II; lane 4, pUCSS04 digested with *Sca*I.

centrifuged (8000 *g* for 60 min) and the supernatant kept as a crude extract (Fraction I). Fraction I was incubated at 85°C for 10 min and then centrifuged (8000 *g* for 60 min) to remove denatured proteins. The clear supernatant (Fraction II) was brought to 70% saturation with ammonium sulfate and stored at 4°C overnight. The precipitate was collected by centrifugation (8000 *g* for 60 min) and dissolved in the same buffer A. This crude enzyme sample was heat-treated at 85°C for 10 min. A heparin column (1.6 × 2.5 cm, HiTrap; Pharmacia, Uppsala, Sweden) was equilibrated with buffer A and the enzyme sample was applied to the column. Protein fractions were eluted with a linear gradient of 0.1–2 M NaCl. KOD DNA polymerase and an intein were recovered at ~0.6–0.7 and 1.1 M NaCl, respectively. The intein fraction (Fraction III) was dialyzed overnight against buffer C (10 mM Na-phosphate, pH 7.0, 0.5 M NaCl, 0.1 mM EDTA, 1 mM DTT) and then applied to the same heparin column equilibrated with buffer C. The intein fraction (Fraction IV) was eluted with a linear gradient of 0.5–2 M NaCl at a flow rate of 1 ml/min. Following elution, the intein fraction was also dialyzed overnight against buffer D (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT) and then concentrated with polyethylene glycol 20 000. Purified protein fractions were concentrated and applied to a Superose 6 HR10/30 gel filtration column (Pharmacia) equilibrated in buffer D at a flow rate of 0.5 ml/min in order to determine the molecular masses. Purity of proteins was examined by 0.1% SDS–12% PAGE.

Amino acid sequence analysis

Purified PI-*Pko*I and PI-*Pko*II were subjected to electrophoresis on a 13% SDS–polyacrylamide gel with a Tris–Tricine buffer system (22) and transferred to a 0.2 µm Trans-Blot Transfer Medium PVDF Membrane (BioRad Labolatories, Hercules, CA) by electroblotting. The membrane was stained with Coomassie blue R-250 and two protein bands of 41 and 62 kDa, corresponding to PI-*Pko*I and PI-*Pko*II, respectively, were excised and each subjected to sequential Edman degradation using an Applied Biosystems ABI 473A-3. The data were acquired and analyzed on an Applied Biosystems 610A Data System.

Enzyme assays

The endonuclease activities of PI-*Pko*I and PI-*Pko*II were examined by digestion of a plasmid which contained the part of the mature DNA polymerase sequence without intein DNA sequences. The 449 bp *Stu*I–*Sac*I fragment of the mature DNA polymerase gene containing both cleavage sites for the two intein endonucleases was cloned into pUC18 and the plasmid named pUCSS04. pUCSS04 (0.5 µg) was linearized with *Sca*I and was then incubated with PI-*Pko*I or PI-*Pko*II at 70°C for 1 h in the standard enzyme reaction buffer containing 50 mM Tris–HCl (pH 8.5), 10 mM $MgCl_2$, 1 mM DTT, 100 mM NaCl. After phenol–chloroform extraction and ethanol precipitation, samples were analyzed by agarose gel electrophoresis. One unit of endonuclease activity was defined as the amount of enzyme required to hydrolyze 1 µg linearized pUCSS04 plasmid completely in 1 h at 70°C.

Determination of cohesive termini generated by endonuclease treatments

T4 DNA polymerase is useful for the examination of cleavage patterns produced by endonucleases owing to its ability to catalyze repair of 5′-overhangs and hydrolysis of 3′-overhangs. pUCSS04 digested with PI-*Pko*I and PI-*Pko*II, respectively, was recovered from the agarose gel with a Genclean II kit (Bio101 Inc., Vista, CA) and the DNA was treated with T4 DNA polymerase (Takara Shuzo Co., Kyoto, Japan) in the presence of dNTPs. The DNA mixture was

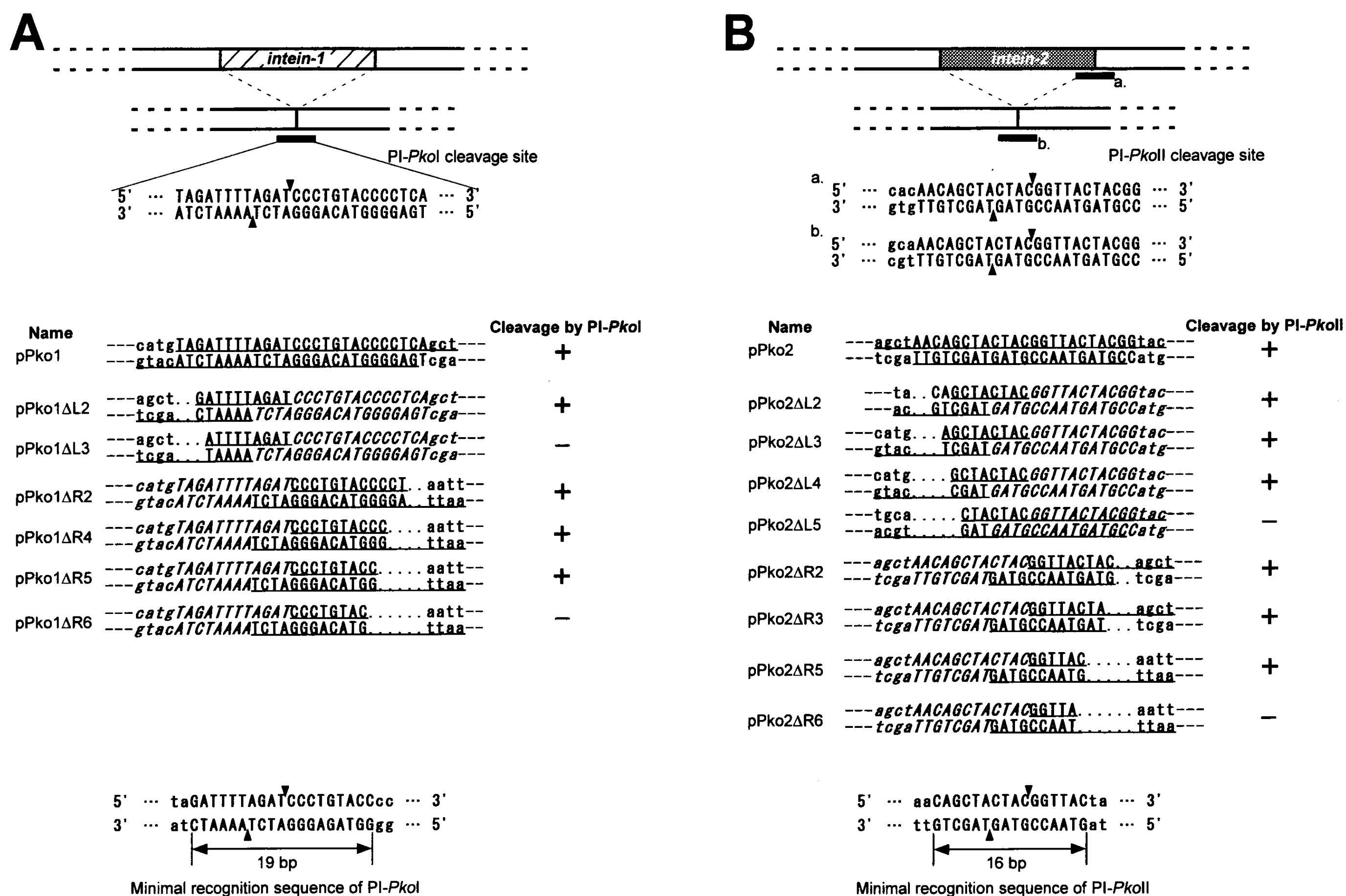

Figure 4. The cleavage sites and the minimal recognition sequences of PI-*Pko*I (**A**) and PI-*Pko*II (**B**). All sequences synthesized, cleavage results and resultant plasmid names are shown in the figure. Underlined sequences indicate the synthesized oligonucleotides. Upper case characters are the nucleotides within recognized sequences and lower case ones are the bases within restriction enzyme sites at the MCS. Italics mark the sequences that are not changed in construction of the plasmids.

then treated with T4 DNA ligase and used to transform *E.coli* cells. Nucleotide sequence analysis was used to determine the nature of cohesive termini generated by endonuclease treatments.

Determination of minimal recognition sequences

pUC19 plasmid containing short oligonucleotides that can be recognized and cleaved by either PI-*Pko*I or PI-*Pko*II was prepared. Then the length of the oligonucleotides was gradually shortened from either end to specify the minimum recognition sequence required for cleavage by each of the intein endonucleases.

RESULTS AND DISCUSSION

Purification and endonuclease activities of inteins

When expression of the KOD DNA polymerase gene containing each of two intein sequences was induced, both inteins and KOD DNA polymerase were recovered as mature forms from the precursor of the KOD DNA polymerase. This implies that protein splicing of KOD DNA polymerase precursors can occur in *E.coli* even at 37°C, even though the gene is derived from the hyperthermophilic archaea. Two intein endonucleases were purified to homogeneity by the method explained above (Fig. 1). Elution profiles of both inteins from the gel filtration chromatography showed that these enzymes were both monomeric. As shown in Figure 2, the N-terminal sequence of purified PI-*Pko*I was (Cys)-His-Pro-Ala and that of PI-*Pko*II was Ser-Ile-Leu-Pro-Glu-Glu, which are identical to the predicted splicing sites based on the conserved protein splicing motif (23). Some intron-encoded proteins, termed inteins, are known to have site-specific DNA endonuclease activity. These intein endonucleases recognize and cleave their respective intein-less DNA alleles. Indeed, KOD1 PI-*Pko*I and PI-*Pko*II could cleave the DNA regions of the respective intein-less DNA sequences. DNA fragment sizes generated by cleavage of pUCSS04 by *Sca*I and PI-*Pko*I and by *Sca*I and PI-*Pko*II were 0.7 and 1.0 kb, respectively, corresponding to the calculated sizes from the DNA sequence (Fig. 3). The nature of the cohesive ends generated by endonuclease treatment was revealed by T4 DNA polymerase treatment followed by ligation and nucleotide sequencing. The cleavage pattern of both PI-*Pko*I and PI-*Pko*II is a 3′ protruding four base overhanging sequence (Fig. 4).

Effects of ion concentrations on endonuclease activities

Purified recombinant PI-*Pko*I and PI-*Pko*II showed superior thermostability. In particular, PI-*Pko*I showed no decrease in endonuclease activity even after 1 h incubation at 90°C. The

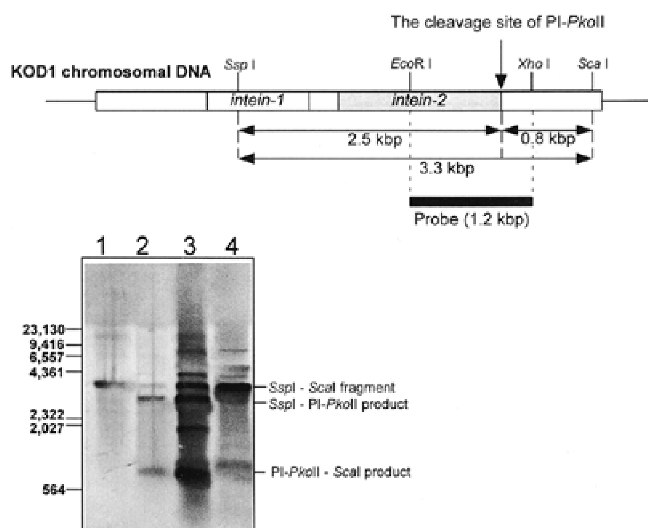

Figure 5. Cleavage of KOD1 chromosomal DNA by PI-*Pko*II. KOD1 chromosomal DNA and pET-*pol* plasmid DNA (positive control) were respectively digested with *Ssp*I and *Sca*I before treatment with PI-*Pko*II. Digested samples were separated by 1% agarose gel electrophoresis and transferred to Hybond N+ membrane (Amersham, Arlington Heights, IL). An *Eco*RI–*Xho*I fragment (1.2 kb) labeled with digoxigenin was hybridized to the membrane and then detected using the DIG detection system (Boehringer Mannheim, Indianapolis, IN). Lane 1, KOD1 chromosomal DNA digested with *Ssp*I and *Sca*I; lane 2, KOD1 chromosomal DNA digested with *Ssp*I, *Sca*I and PI-*Pko*II; lane 3, pET-*pol* digested with *Ssp*I, *Sca*I and PI-*Pko*II; lane 4, pET-*pol* digested with *Ssp*I and *Sca*I. DNA sizes are shown on the left. The *Ssp*I–*Sca*I (3.3 kb), *Ssp*I–PI-*Pko*II (2.5 kb) and PI-*Pko*II–*Sca*I (0.8 kb) fragments are also indicated.

specific activities of these two endonucleases were examined by a dilution method as reported before (24). PI-*Pko*I (17 700 U/mg) exhibited 22 times higher specific activity than PI-*Pko*II (810 U/mg). The effects of NaCl and KCl concentration on endonuclease activity of PI-*Pko*I and PI-*Pko*II were also examined. Both enzymes had higher activity with potassium ions than sodium ions. The endonuclease activity of PI-*Pko*I was lower in 0.5 M NaCl or KCl and could not be detected in 1 M NaCl or KCl. The activity of PI-*Pko*II was detectable at 0.75 M NaCl but was not detectable at 1 M NaCl, although the enzyme was still active at 1 M KCl. The concentration of intracellular potassium ions is very high (>0.5 M) in the cytoplasm of hyperthermophiles (25) and, therefore, the effects of ionic strength on the activity of both endonucleases suggest that PI-*Pko*II is probably very active in KOD1 cells while PI-*Pko*I might be less active.

Minimal recognition sequences of the endonucleases

Recognition sequences of only a few homing endonucleases have been previously reported (18,20,26). In hyperthermophiles, three inteins encoded in DNA polymerase genes are known as homing endonucleases. However, their recognition sequences have not been precisely determined (12). Minimal recognition sequences cleaved by PI-*Pko*I and PI-*Pko*II were determined. All sequences of oligonucleotides used and the results of digestion by each endonuclease are summarized in Figure 4. The minimal sequences for the intein endonucleases are 19 bp (5′-GATTTTAGATCCCTG-TACC-3′) for PI-*Pko*I and 16 bp (5′-CAGCTACTACGGTTAC-3′) for PI-*Pko*II (Fig. 4).

Cleavage of chromosomal DNA by PI-*Pko*II

As mentioned above, PI-*Pko*I and PI-*Pko*II cleave the DNA in the region of the intein-less allele. However, we found the recognition and cleavage site of PI-*Pko*II at the downstream junction between intein (PI-*Pko*II) and mature KOD DNA polymerase. Cleavage by PI-*Pko*II at this position could be observed even when KOD1 chromosomal DNA was used (Fig. 5).

Since KOD1 cells survive in spite of expression of PI-*Pko*II, it is suggested that some mechanism must exist to protect KOD1 chromosomal DNA from digestion by PI-*Pko*II. We speculate that these intein endonucleases may play a role in chromosomal DNA rearrangement, as has been reported for the intein endonuclease from *S.cerevisiae* (14). Further studies will be necessary to elucidate a protection mechanism against digestion and the relationship of PI-*Pko*I and PI-*Pko*II to intein mobilization and DNA rearrangement.

ACKNOWLEDGEMENT

This work was supported by a grant from CREST (Core Research for Evolutional Science and Technology).

REFERENCES

1 Chong,S., Shao,Y., Paulus,H., Benner,J., Perler,F.B. and Xu,M.-Q. (1996) *J. Biol. Chem.*, **271**, 22159–22168.
2 Cooper,A.A., Chen,Y., Lindorfer,M.A. and Stevens,T.H. (1993) *EMBO J.*, **12**, 2575–2583.
3 Cooper,A.A. and Stevens,T.H. (1995) *Trends Biochem. Sci.*, **20**, 351–356.
4 Xu,M.-Q., Comb,D., Paulus,G.H., Noren,C.J., Shao,Y. and Perler,F.B. (1994) *EMBO J.*, **13**, 5517–5522.
5 Xu,M.-Q., Southworth,M.W., Mersha,F.B., Hornstra,L.J. and Perler,F.B. (1993) *Cell*, **75**, 1371–1377.
6 Hirata,R., Ohsumi,Y., Nakano,A., Kawasaki,H., Suzuki,K. and Anraku,Y. (1990) *J. Biol. Chem.*, **265**, 6726–6733.
7 Kane,P.M., Yamashiro,C.T., Wolczyk,D.F., Neff,N., Goebel,M. and Stevens,T.H. (1990) *Science*, **250**, 651–657.
8 Davis,E.O., Jenner,P.J., Brooks,P.C., Colston,M.J. and Sedgwick,S.G. (1992) *Cell*, **71**, 201–210.
9 Davis,E.O., Thangaraj,H.S., Brooks,P.C. and Colston,M.J. (1994) *EMBO J.*, **13**, 699–703.
10 Hodgee,R.A., Perler,F.B., Noren,C.J. and Jack,W.E. (1992) *Nucleic Acids Res.*, **20**, 6153–6157.
11 Perler,F.B., Comb,D.G., Jack,W.E., Moran,L.S., Qiang,B., Kucera,R.B., Benner,J., Slatko,B.E., Nwankwo,D.O., Hempstead,S.K., Carlow,C.K.S. and Jannasch,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5577–5581.
12 Perler,F.B., Olsen,G.J. and Adam,E. (1997) *Nucleic Acids Res.*, **25**, 1087–1093.
13 Gimble,F.S. and Stephens,B.W. (1995) *J. Biol. Chem.*, **270**, 5849–5856.
14 Gimble,F.S. and Thorner,J. (1992) *Nature*, **357**, 301–306.
15 Gimble,F.S. and Thorner,J. (1993) *J. Biol. Chem.*, **268**, 21844–21853.
16 Gimble,F.S. and Wang,J. (1996) *J. Mol. Biol.*, **263**, 163–180.
17 Belfort,M. and Roberts,R.J. (1997) *Nucleic Acids Res.*, **25**, 3379–3388.
18 Dalgaard,J.Z., Garrett,R.A. and Belfort,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5414–5417.
19 Lambowitz,A.M. and Belfort,M. (1993) *Annu. Rev. Biochem.*, **62**, 587–622.
20 Lykke-Andersen,J., Thi-Ngoc,H.P. and Garrett,R.A. (1994) *Nucleic Acids Res.*, **22**, 4583–4590.
21 Takagi,M., Nishioka,M., Kakihara,H., Kitabayashi,M., Inoue,H., Kawakami,B., Oka,M. and Imanaka,T. (1997) *Appl. Environ. Microbiol.*, **63**, 4504–4510.
22 Schagger,H. and von Jagow,G. (1987) *Anal. Biochem.*, **166**, 368–379.
23 Pietrokovski,S. (1994) *Protein Sci.*, **3**, 2340–2350.
24 Halford,S.E., Taylor,J.D., Vermote,C.L.M. and Vipond,I.B. (1994) In Kneale,G.G. (ed.) *DNA–Protein Interactions. Methods in Molecular Biology.* Humana Press, Totowa, NJ, Vol. 30, pp. 385–396.
25 Hensel,R. and Konig,H. (1988) *FEMS Microbiol. Lett.*, **49**, 75–79.
26 Aagaard,C., Awayes,M.J. and Garrett,R.A. (1997) *Nucleic Acids Res.*, **25**, 1523–1530.