A putative Type IIS restriction endonuclease GeoICI from *Geobacillus* sp. – A robust, thermostable alternative to mezophilic prototype BbvI

JOANNA ZEBROWSKA[1], OLGA ZOŁNIERKIEWICZ[1], MARTA A SKOWRON[2], AGNIESZKA ZYLICZ-STACHULA[1], JOANNA JEZEWSKA-FRACKOWIAK[1] and PIOTR M SKOWRON[1,*]

[1]*Department of Molecular Biotechnology, Institute for Environment and Human Health Protection, Faculty of Chemistry, and* [2] *Department of Molecular Biology, Faculty of Biology, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland*

**Corresponding author (Email, piotr.skowron@ug.edu.pl; pmars44@wp.pl)*

Screening of extreme environments in search for novel microorganisms may lead to the discovery of robust enzymes with either new substrate specificities or thermostable equivalents of those already found in mesophiles, better suited for biotechnology applications. Isolates from Iceland geysers' biofilms, exposed to a broad range of temperatures, from ambient to close to water boiling point, were analysed for the presence of DNA-interacting proteins, including restriction endonucleases (REases). GeoICI, a member of atypical Type IIS REases, is the most thermostable isoschizomer of the prototype BbvI, recognizing/cleaving 5′-GCAGC($N_{8/12}$)-3′ DNA sequences. As opposed to the unstable prototype, which cleaves DNA at 30°C, GeoICI is highly active at elevated temperatures, up to 73°C and over a very wide salt concentration range. Recognition/cleavage sites were determined by: (i) digestion of plasmid and bacteriophage lambda DNA (λ); (ii) cleavage of custom PCR substrates, (iii) run-off sequencing of GeoICI cleavage products and (iv) shotgun cloning and sequencing of λ DNA fragmented with GeoICI. *Geobacillus* sp. genomic DNA was PCR-screened for the presence of other specialized REases-MTases and as a result, another putative REase-MTase, GeoICII, related to the *Thermus* sp. family of bifunctional REases-methyltransferases (MTases) was detected.

[Zebrowska J, Zolnierkiewicz O, Skowron MA, Zylicz-Stachula A, Jezewska-Frackowiak J and Skowron PM 2016 A putative Type IIS restriction endonuclease GeoICI from *Geobacillus* sp. – A robust, thermostable alternative to mezophilic prototype BbvI. *J. Biosci.* **41** 27–38] **DOI** 10.1007/s12038-016-9595-z

1. Introduction

Soil from geothermal locations harbours a wide variety of thermophilic microorganisms, which comprise a rich source of compounds for industrial and scientific applications. These include secondary metabolites, antimicrobial compounds and waste utilization enzymes, such as REases for DNA manipulation purposes, among others. In a bacterial cell, REases are a part of the restriction-modification system (R-M), which typically acts as a functional equivalent of *Eucaryotes*' immune system, protecting the cell from an invasion of foreign DNA, e.g. a bacteriophage attempting to infect the bacterium or through free DNA uptake *via* transformation or transfection.

The R-M systems have been classified into four major types basing on their genetics and the organization of their functional domains, cofactor requirements, as well as their modes of recognition and cleavage. Out of the 4192 currently known, biochemically or genetically characterized REases (Roberts *et al.* 2015; *http://rebase.neb.com/rebase/statlist.html*), 4039 belong to the Type II R-M systems, which together with 15 371 putatives comprise the largest biochemically characterized fraction. Only 464 of these REases (all types: *http://rebase.neb.com/cgi-bin/seqspelist*), however,

Keywords. *Geobacillus*; restriction endonuclease; subtype IIS, IIC, IIG; thermostability

Published online: 15 February 2016

are prototype enzymes with unique DNA recognition specificities, whereas remaining are isoschizomers (Roberts *et al.* 2015). Subtype IIS enzymes atypical REases that recognize a specific DNA sequence and cleave at a defined distance from this sequence, up to 21 nucleotides (nt) away. This is a small group of REases, however new enzymes are constantly being discovered. The majority of known Type II REases have been isolated from mesophilic bacteria, and thus are stable at temperatures below 45°C. Relatively few REases are capable of retaining their activity at higher temperatures and these enzymes are very useful molecular tools (Roberts *et al.* 2015). Certain thermostable Type IIS/IIG/IIC REases display a novel type of specificity relaxation, towards very frequent DNA cleavage. As opposed to the classic 'star activity' caused by altered reaction conditions (suboptimal pH, salt concentration, high enzyme concentration, organic solvents, high enzyme to substrate ratio), some members of the '*Thermus* family' bifunctional enzymes described by our team (Skowron *et al.* 2003) exhibit a specificity shift caused by S-adenosylmethionine cofactor's analogue – sinefungin. Currently, there are only several natural REases displaying very frequent cleavage (with a recognition site shorter than 4 bp): CviJI/CviJI* (Xia *et al.* 1987; Skowron *et al.* 1995; Mead *et al.* 1995; Gingrich *et al.* 1996; Swaminathan *et al.* 1998), SetI (Roberts *et al.* 2015), FaiI (Roberts *et al.* 2015), BspANI (Roberts *et al.* 2015) and BpuJo21 (Roberts *et al.* 2015). As genomic and metagenomic research depends on the construction of representative libraries, we continue to screen geothermal samples for the presence of very frequently cleaving or atypical Type IIS/IIG/IIC thermostable REases. In this work we investigated samples from Iceland geysers and the biofilms of surrounding soils, where the temperature gradually decreases from 100°C to ambient. The isolated thermophilic bacteria were screened biochemically for the presence of REases, one of the isolates was assigned to the genus *Geobacilllus sp.* and produced GeoICI, a thermostable isoschizomer of the Type IIS REase BbvI from the mesophilic *Bacillus brevis* (Hattman *et al.* 1978; Roberts *et al.* 2015). As BbvI is an unstable enzyme, the presented GeoICI is a more practical tool for gene cloning and analysis.

2. Materials and methods

2.1 *Bacterial strains, plasmids, media and reagents*

The *Geobacillus* sp. strain was originally found around Iceland hot-water geysers. Biofilm samples containing bacteria were streaked on agar plates with a modified *Thermus* medium TM (3% peptone, 2% yeast extract, Castenholz Basal Salt Solution, Nitch's Trace Elements, 2.0% agar (ATCC)) and incubated at various temperatures for 48 h. The dominant strain visible on the agar plate was chosen for further microbiological and molecular characterization, which included colony morphology description, Gram staining, growth temperature optimization, identification through MALDI (BRUSS Group ALAB Sp. z o.o., Gdynia, Poland) mass spectrometry and 16S rRNA sequencing and analysis. The 16S rRNA segment was PCR amplified from *Geobacillus* sp. genomic DNA using universal primers (Yoon *et al.* 1998) and the amplification products were sequenced. The sequence was compared with sequences from the GenBank Database and analysed using BLAST and the online programme MUSCLE (Altschul *et al.* 1997; Edgar 2004).

Escherichia coli (*E. coli*) DH11S electrocompetent cells {*mcrA* Δ[*mrrhsdRMS*(rK-, mK+)- *mcrBC*] Δ(*lac-proAB*) Δ(*recA1398*) *deoR*, *rpsL*, *srl-thi*, *supE*/*F' proAB*+ *lacI*QZΔ*M15*} were from Life Technologies (Gaithersburg, MD, USA). Media components were from Scharlab S.L., Barcelona, Spain. Agarose, Heparin-agarose were from Vivantis (Subang Jaya, Malaysia), DEAE-cellulose and Phosphocellulose P11 were from Whatman (Springfield Mill, UK). All other reagents were from Sigma-Aldrich (St. Louis, MO, USA) or Amresco (Solon, OH, USA) of the highest available purity. Plasmids: pUC19, pBR322, pACYC184, pUCmini (modified pUC19, supplementary file 1), pThermus, were from Piotr Skowron's collection. REases, T4 DNA Ligase, T4 DNA Polymerase, λ DNA, 100 bp and 1 kb DNA ladders were from Thermo Fisher Scientific Baltics UAB (Waltham, MA, USA). Miniprep DNA isolation kits, DNA purification kits and thermostable proofreading Marathon DNA polymerase were from A&A Biotechnology (Gdynia, Poland). The protein Ladder was from GE Healthcare (Little Chalfont, United Kingdom). PCR primer synthesis and DNA sequencing were performed at Genomed (Warsaw, Poland).

2.2 *GeoICI REase isolation from Geobacillus sp.*

The bacterial lysate was prepared as follows: a single bacterial colony was inoculated into 1.5 mL of TM liquid medium, and used to inoculate 10 L of TM medium in a New Brunswick Scientific BioFlo 115 bioreactor under aerobic conditions at 67.5°C, subsequently the culture was chilled down and centrifuged. The bacterial pellet was suspended in DEAE-cellulose buffer A (50 mM Tris-HCl pH 8.0, at 25°C, 0.5 mM EDTA, 50 mM NaCl, 0.05% Triton X-100, 5 mM β-mercaptoethanol (β-ME), 1 mM PMSF). Lysozyme was added to a final concentration of 0.5 mg/mL and the suspension was incubated on ice for 1 h, sonicated and centrifuged. Nucleic acids and acidic proteins were removed by the addition of polyethylenimine to 0.5% and subsequent stirring at 4°C for 30 min. Insoluble complexes were removed by centrifugation. The obtained supernatant was subjected to ammonium sulphate protein fractionation, the 35–65% saturation fraction was centrifuged, dissolved in buffer A and extensively dialysed against buffer A. Proteins were applied

to DEAE-cellulose and washed with buffer A, GeoCI was eluted in several steps with a growing concentration of NaCl, within the range of 100–400 mM. The fractions containing GeoICI activity were combined and dialysed against phosphocellulose buffer B (20 mM K/PO_4 pH 8.0, at 25°C, 1 mM EDTA, 50 mM NaCl, 50 mM NaCl, 0.01% Triton-X-100, 5% glycerol, 5mM β-ME, 0.5 mM PMSF) and applied to a Phosphocellulose P11 column, washed with buffer B and GeoICI was eluted in several steps with a 100-800 mM NaCl concentration. Fractions containing GeoICI were pooled and dialysed against heparin-agarose buffer C (20 mM Tris-HCl pH 9.5, at 25°C, 1 mM EDTA, 50 mM NaCl, 0.01% Triton X-100, 5% glycerol, 5 mM β-ME, 0.5 mM PMSF). The column was washed with buffer C and GeoICI was eluted with a series of steps with an NaCl concentration within the range of 100-500 mM. Fractions, containing GeoICI were pooled and dialysed against storage buffer S (20 mM Tris-HCl pH 8.0, at 25°C, 1 mM EDTA, 200 mM NaCl, 5 mM ditiotreitol, 0.01% Tween-100, 0.02% Triton-X-100, 50% glicerol, 0.5 mM PMSF). GeoICI activity in crude lysate and chromatography fractions were assayed by incubation with pUCmini DNA for 1 h at 65°C in 25 μL reactions containing 500 ng of plasmid DNA in medium salt buffer (10 mM Tris-HCl pH 7.5 at 50°C, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol, 100 μg/mL bovine serum albumin) and analysed through 1.3% agarose gel electrophoresis in 0.5x TBE buffer, followed by staining with ethidium bromide. Because many Type IIS REases do not cleave substrate DNA completely, one unit of GeoICI is defined as the amount of enzyme required to hydrolyse 1 μg of λ DNA in 1 h at 50°C in 50 μL of standard GeoICI buffer resulting in a complete or stable partial DNA cleavage pattern.

2.3 *Determination of GeoICI recognition and cleavage sites*

Using a purified GeoICI preparation, the recognition site and cleavage sites were determined by: (i) digestion of plasmid and λ DNA; (ii) cleavage of short PCR substrates, (iii) run-off sequencing of GeoICI cleavage products; and (iv) shotgun cloning and sequencing of cleaved λ DNA.

In approach (i), 0.5 μg of plasmid pUC19, pUCmini and λ DNA were digested for 1 h with GeoICI and analysed through 1.3% agarose gel electrophoresis. Next, the putative GeoICI recognition site was confirmed by additional mapping of doubly digested pUCmini DNA using BsaI, ScaI or XmaI (data not shown). To simplify digestion patterns interpretation, prior to digestion by the second reference REase, GeoICI digestion products were subjected to PAGE and selected bands were isolated using electroelution, phenol/chloroform extraction and ethanol precipitation. Digestion patterns obtained in all experiments were analysed using DNASIS MAX software (Hitachi Software, San Bruno, CA, USA) as well as REBASE Tools: REBsites, NEBcutter (Vincze *et al.* 2003), and REBpredictor (Gingeras *et al.* 1978; Roberts *et al.* 2015). In approach (ii) recognition and cleavage sites were verified using custom PCR substrates with isolated putative GeoICI recognition sites. Four substrates were used: (1) 377-bp PCR fragment amplified from shotgun clone no 5, containing a λ DNA segment from the GeoICI digestion products library prepared in pUC19. This substrate contained the putative recognition sequence 5′-GCAGC-3′ and was amplified with primers: 5′-CGCCAGGGTTTTCCCAGTCACG-3′ and 5′-AGCGGATAACAATTTCACACAGG-3′.

The PCR reactions were performed in 100 μL volumes and contained: 1x Marathon PCR Buffer (contains 5% DMSO), 0.4 mM dNTPs, 0.5 μM of each primer, 1 ng DNA from the clone and 1 unit of Marathon DNA Polymerase, the cycling profile was as follows: 95°C for 3 min, 89°C for 20 s (addition of DNA polymerase), 30 cycles of 94°C for 30 s, 60°C for 30 s, and 68°C for 30 s and a final extension at 68°C for 2 min. The analogous 204-bp PCR substrate was amplified from clone no 21 using the same cycling profile with the following primers: 5′-CGCCAGGGTTTTCCCAGTCACG-3′ and 5′-AGCGGATAACAATTTCACACAGG-3′. The 198-bp substrate was amplified from pACYC184 with the primers: 5′-CCTTCAACCCAGTCAGCTCCT-3′ and 5′-CGAGGGCGTGCAAGATTCC-3′, under the same conditions, except for the cycling profile which was as follows: 97°C for 3 min, 89°C for 20 s (addition of DNA polymerase), 30 cycles of 94°C for 30 s, 55°C for 30 s, 68°C for 30 s and a final extension step of 68°C for 2 min. The 199-bp substrate was amplified from pBR322 with the primers: 5′-CATGATCGTGCTCCTGTCGT3′ and 5′-GGTGCAGGGCGCTGACTTCC-3′, under the same conditions, except for the cycling profile which was as follows: 96°C for 3 min, 89°C for 20 s (addition of DNA polymerase), 30 cycles of 94°C for 30 s, 50°C for 30 s, 70°C for 30 s and a final extension at 70°C for 2 min. PCR products were purified using a DNA purification kit. GeoICI digestions of purified PCR products were carried out in 50 μL reaction volumes at 50°C for 2 h. When the reactions were complete, samples were purified by phenol/chloroform extraction and ethanol precipitated, the products were analysed through PAGE. As Type IIS REases cleave outside their recognition sequences, the recognition/cleavage sites were confirmed in the biochemical approach (iii), where a GeoICI-digested 198-bp substrate DNA, pUCmini and a 1789-bp PCR fragment of pACYC184 were used for run-off sequencing. The 1789-bp PCR substrate was obtained from pACYC184 using the following primers: 5′-CATCAGCGCTAGCGGAGTGTA-3′ and 5′-CGAGGGCGTGCAAGATTCC-3′. Amplification was performed in 100 μL volumes and contained: 1x Marathon PCR Buffer, 0.4 mM dNTPs, 0.5 μM of each primer, 15 ng pACYC184 DNA, 1.5 units of Marathon DNA Polymerase. The PCR cycling profile was: 97°C for 4 min, 89°C for 20 s (addition of DNA polymerase), 94°C for 30 s, 55°C for 30 s, 68°C for 2 min (for 30

Figure 1. Gram staining of *Geobacillus* sp. bacteria isolated from geysers biofilm samples. Preparation was analysed under light microscope at 1250× magnification.

cycles) and a final extension at 68°C for 1.5 min. The GeoICI digestion products were treated with proteinase K to remove DNA-binding proteins, phenol/chloroform extracted and ethanol precipitated, followed by PAGE separation and gel isolation. Purified DNA was subjected to run-off Sanger automated sequencing (Genomed, Warsaw, Poland). For pUCmini-derived PCR substrates the following primers were used: 5′-GCGACACGGAAATGTTGAATAC-3′ and 5′-GGTGAA-GATCCTTTTTGATAATCTCACGACCAAAATCCCT-TAACGTGAG-3′, for 198-bp PCR substrate: 5′-CCTTCAACCCAGTCAGCTCCT-3′ and 5′-CGAGGGCGTGCAAGATTCC-3′, for the 1789-bp PCR substrate: 5′-CATCAGCGCTAGCGGAGTGTA-3′ and 5′-CGAGGGCGTGCAAGATTCC-3′. Sequencing data was analysed using Chromas Lite 2.1.1. software (Technelysium Pty Ltd., South Brisbane, Australia) and DNASIS MAX (Hitachi Software, San Bruno, CA, USA). Approach (iv), based on shotgun cloning and sequencing of the GeoICI, digested 48,500-bp λ DNA was used to confirm previous results as well as to evaluate the potential presence of cognate site variants, which could have been absent in relatively short substrates. 1.5 µg of λ DNA was digested with 2 units of GeoICI at 50°C

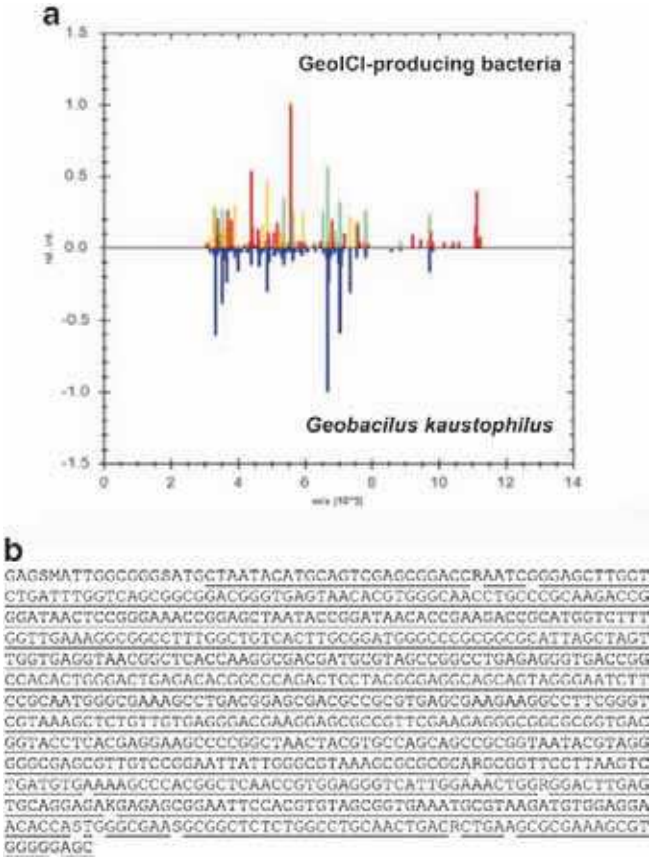


Figure 2. Peptide profile identification of *Geobacillus* genus by MALDI mass spectrometry method. (**a**)Top panel shows peptide profile of investigated bacterial sample, encoding GeoICI. Bottom panel shows peptide profile of *Geobacilus kaustophilus* from Database NCBI Identifier. Green colour indicates highly probable species identification status, yellow – probable species identification status, red – not reliable identification status. (**b**) 762-bp nt sequence of 16S rRNA gene fragment PCR amplified from bacteria producing GeoICI. Nt homologous to 16S rRNA of *Geobacilus kaustophilus* are underlined.

for 3 h. Reaction products were blunted with T4 DNA polymerase/dNTPs, separated through PAGE, electroeluted, phenol chloroform/chloroform extracted and ethanol precipitated and cloned into the SmaI site of pUC19. The resulting ligation mixture was transformed into *E. coli* DH11S cells, cultured on X-Gal/IPTG/ampicillin plates and blue/white screened through α-complementation (Green and Sambrook 2012). White colonies were used directly as templates in PCR with pUC19 primers: 5′-CGCCAGGGTTTTCCCAGTCACGACG-3′ and 5′-AGCGGATAACAATTTCACACAGG-3′. The PCR reactions were performed in 20 µL volumes and contained: 1x Marathon PCR Buffer, 0.4 mM dNTPs, 0.5 µM of each primer and 1 unit of Marathon DNA Polymerase. The PCR cycling profile was as follows: 95°C for 3 min, 30 cycles of 94°C for 30 s, 60°C for 30 s, 68°C for 30 s and a final extension at 68°C for 2 min. Bacteria containing recombinant plasmids with inserts were cultured in LB medium, plasmids were isolated using a DNA isolation kit and sequenced with primers flanking the MCS of pUC19.

2.4 *Characterization of GeoICI*

Evaluations of reaction optima, such as temperature, pH and salt concentrations were determined using 300 ng of the pThermus plasmid (8705 bp, GC% 58,07), and 300 ng of the 1789-bp PCR substrate. The reaction temperature optimum was determined in medium salt buffer, which had been initially shown to yield best digestions. The NaCl concentration optimum was also determined in this buffer, devoid of initial salt content. The pH optimum was evaluated in three buffer systems, each dedicated to the maximum buffering capacity range: sodium acetate-acetic acid with a pH (50°C) from 4.0 to 5.5, HEPES-KOH pH from 6.0 to 7.0, and Tris–HCl buffer pH from 7.5 to 9.0 (Zylicz-Stachula *et al*. 2011a, b). The pH of the reaction buffers was adjusted at 50°C after all the buffer components had been dissolved. Cleavage reactions were performed for 1 h using limiting amounts of GeoICI.

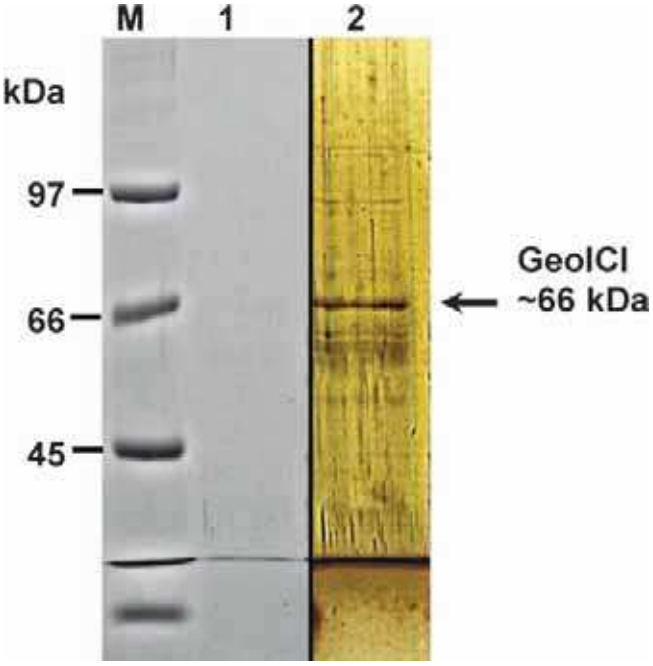


Figure 3. Purification of GeoICI REase. 10% SDS/PAGE of purified enzyme. Lane M, protein marker (GE Healthcare); lane 1, purified preparation of GeoICI, Coomassie blue stained; lane 2, purified preparation of GeoICI, silver stained (Green and Sambrook 2012).

2.5 *Evaluation of the presence of genes coding for bifunctional Thermus sp. family enzymes in Geobacillus sp. genomic DNA*

In addition to the major aim of this work, the *Geobacillus* sp. was screened for the presence of genes coding for Type IIS/IIG/IIC family of atypical, bifunctional *Thermus* sp. REases, which our team had defined in previous papers (Skowron *et al.* 2003; Krefft *et al.* 2015). For screening purposes, a 450-bp segment, located within a conserved MTase domain, was selected as a diagnostic marker (Zylicz-Stachula *et al.* 2009; Skowron *et al.* 2013; Krefft *et al.* 2015). A set of the following PCR primers were used: 5′-GTTACCGTGATAGACCCGGCCATGGGCACGGGGACCTT-3′ and: 5′-CACCCGGTCGTAGGGGGGGGCGCCAAGGATCACCA-3′. The 50 µL reactions contained: 1x Marathon

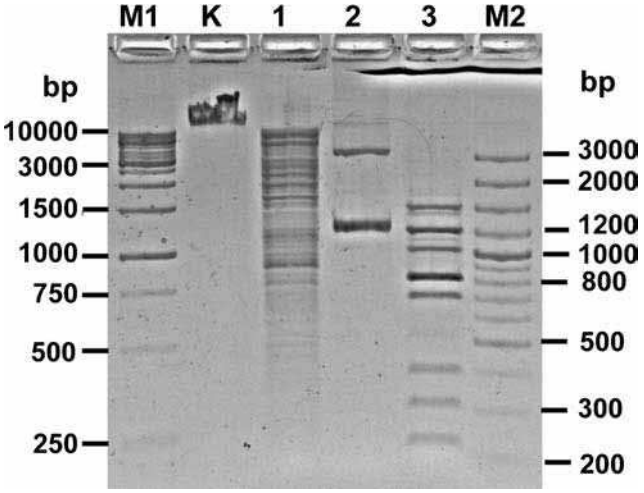


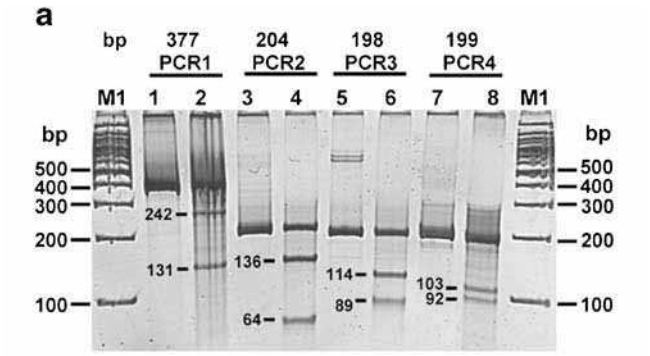
Figure 4. GeoICI digestion of standard DNA substrates. Agarose gel electrophoresis-analysed digestions of standard substrates: Lane M1, GeneRuler™ 100- bp DNA Ladder (selected bands marked); lane K, undigested λ DNA; lane 1, GeoICI-digested λ DNA; lane 2, undigested pUCmini DNA; lane 3, GeoICI-digested pUCmini DNA; lane M2, GeneRuler™ 1 kb DNA Ladder (selected bands marked). The size estimation error was approximately 5%.

PCR Buffer, 0.4 mM dNTPs, 0.5 μM of each primer, 20 ng genomic DNA of either *Geobacillus* sp. or *Thermus* sp. as a control and 1 unit of Marathon DNA Polymerase. The profile was as follows: 97°C for 4 min, 89°C for 20 s (addition of DNA polymerase), 30 cycles of 95°C for 30 s, 69°C for 30 s, 72°C for 20 s and a final extension at 69°C for 1 min.

3. Results and discussion

3.1 *Geobacillus sp. isolation and GeoICI purification*

Geobacillus sp. bacteria were originally found in biofilms surrounding Iceland geysers, where the water temperature reached 100°C and then gradually decreased when flowing



b

GeoICI site 5'-GCAGC-3'

PCR1 **377 bp**

5'ACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGA
GCTCGGTACCCTTATCTGGTTGCCGACGGATGGTGATGCCGAGAACTTTATGAAA
ACCCACGTTGAGCCGACTATTCGTGATATTCCGTC**GCTGC**TGGCGCTGGCCCCGT
GGTATGGCAAAAAGCACCGGGATAACACGCTCACCATGAAGCGTTTCACTAATG
GGCGTGGCTTCTGGTGCCTGGGCGGTAAAGCGGCAAAAAACTACCGTGAAAAGT
CGGTGGATGTGGCGGGTTATGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCT3'

PCR2 **204 bp**

5'ACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGA
GCTCGGTACCCACTATCAAATAT**GCTGC**TTGAGGCTTATTCGGGCGCAGATCTGA
CCAAGCGACAGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTA
ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCT3'

PCR3 **198 bp**

5'CCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCG
CACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCG**GCAGC**GCT
CTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTG
TCGCTTGCGGTATTCGGAATCTTGCACGCCCTCG3'

PCR4 **199 bp**

5'CATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTAC
TGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACT**GCTGCTG**
CAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTGGGTGTTTCG
TAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACC3'

Figure 5. GeoICI cleavage pattern on PCR products, containing isolated putative recognition site. (**a**) GeoICI digestion of PCR substrates. Reaction products were electrophoresed on 15% PAGE/TBE. Lane M, GeneRuler™ 100-bp Ladder (selected bands marked); lane 1, undigested 377-bp PCR substrate; lane 2, GeoICI digested 377-bp PCR substrate; lane 3, undigested 204-bp PCR substrate; lane 4, digested 204-bp PCR substrate; lane 5, undigested 198-bp PCR substrate; lane 6, GeoICI-digested PCR substrate; lane 7, undigested 199-bp PCR substrate; lane 8, digested 199-bp PCR substrate. Lane M, GeneRuler™ 100 bp Ladder. (**b**) PCR substrates sequences. Putative recognition sequences of GeoICI are marked in bold and underlined. Arrows mark GeoICI cleavage sites.

out to geothermal soil, thus it was expected that the enzyme would show a wide temperature activity range. Bacteria formed white, small and round colonies on agar plates, while older colonies (over 48 h cultivation), produced small amounts of an orange pigment. Growth was observed in the temperature range of 40–72°C, with an optimum at 67.5°C. Gram staining resulted in Gram-negative rod-shaped 2–3 μm long cells (figure 1), despite its taxonomic classification as a Gram-positive short rod (Cuebas *et al.* 2011). Endospores were absent in young cultures but in older ones endospore-like structures were visible. Isolated bacteria were analysed through MALDI mass spectrometry, which enables the identification of bacteria basing on the analysis/database comparison of unique protein profiles characteristic for a particular species. Four range values are used, each has an assigned test credibility: 2.300–3.000, highly probable species identification; 2.000–2.299, secure genus identification, but probable species identification; 1.700–1.999, probable genus identification and under 1.700 – the result is not reliable to identification. For bacteria producing GeoICI the value of 1.918 was obtained, showing close similarity to the peptide profile of *Geobacilus kaustophilus* from Database NCBI Identifier, indicating a genus identification status as 'secure', however a more precise species identification status was stated as 'probable' (Fox 2006; Azarko and Wendt 2011a, b). The isolated bacteria were therefore temporarily classified as belonging to the *Geobacillus* genus (figure 2a). Further evaluation was based on the 16S rRNA gene amplification and sequence analysis. The obtained 762 bp PCR product DNA sequence (figure 2b) was blasted through the BLASTn tool indicating that five most closely related species (98%) as of 30.12.2015 are: (i) *Geobacillus kaustophilus* HTA426 (GenBank NC_006510.1); (ii) *Geobacillus thermoleovorans* CCB US3 UF5 (GenBank NC_016593.1; Gupta *et al.* 2012); (iii) *Geobacillus* sp. C56-T3 (GenBank NC_014206.1); (iv) *Geobacillus* sp. Y412MC52 (GenBank NC_014915.1); (v) *Geobacillus* sp. Y412MC61 (GenBank NC_013411.1). Thus, we have shown, using two independent methods, that the bacterial strain, producing GeoICI belongs to *Geobacillus* sp. However, neither the BLASTp evaluation of GenBank annotations nor REBASE Closest Neighbor search engine shows no homology to BbvI coding genes within the above or any other *Geobacillus* sp. This can be a result of remote geographical locations of the listed isolates. Another explanation is that in general, homologies amongst REases are rare, even between isoschizomers. Considering the highly thermophilic nature of GeoICI and mesophilic nature of BbvI, it is likely that their gene sequences are significantly different on the DNA level, since thermophiles in general exhibit a bias towards GC-rich codons, also on the aa level sequences may be different, because thermophiles are under evolutionary pressure to produce more compact proteins cores and rigid structures, which enforces changes in aa sequences and drives divergent evolution.

The presented investigation had two goals: to search for novel DNA cleavage specificities or isoschizomers with improved characteristics for DNA manipulations and to search for genes homologous to the IIS/IIG/IIC family of atypical, bifunctional *Thermus* sp. REase-MTases (Skowron *et al.* 2003; Skowron *et al.* 2013). Currently, this family includes the following members, characterized and cloned in our laboratory: TaqII, TthHB27I, Tth111II, TsoI, TspDTI and TspGWI (Zylicz-Stachula *et al.* 2002, 2009, 2012, 2014; Skowron *et al.* 2013; Jezewska-Frackowiak *et al.* 2015; Krefft *et al.* 2015; Krefft and Skowron, unpublished observations). Genomic DNA of *Geobacillus* sp. was subjected to comparative PCR amplification of the selected 450-bp MTase domain-coding segment, along with a representative of the *Themus* sp. family of enzymes – *Thermus* sp. GW genomic DNA (Skowron *et al.* 2003; Zylicz-Stachula *et al.* 2009). The forward primer was

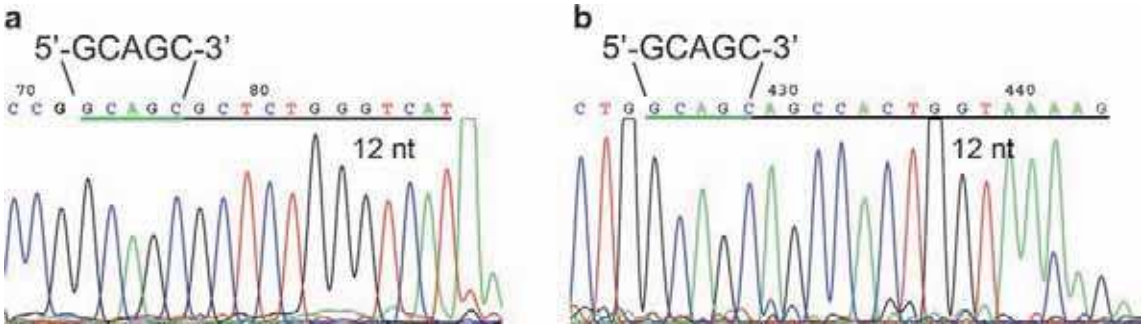


Figure 6. Run-off sequencing determination of GeoICI recognition and cleavage sites. Run-off sequencing chromatograms are showing GeoICI cleavage site in: (**a**) 198-bp PCR substrate sequenced with forward primer 5′-CCTTCAACCCAGTCAGCTCC-3′. (**b**) pUCmini (1698 bp) sequenced with forward primer 5′-GCGACACGGAAATGTTGAATAC-3′ (section 2.3), GeoICI recognition sites boxed and cleavage distance underlined.

Table 1. Determination of GeoICI recognition site by shotgun cloning and sequencing of GeoICI restriction fragments

Sequence position in λ genome or plasmid [bp]	**DNA sequences flanking GeoICI cleavage sites**
38810-38856	5′-*GCGCAGATCTGACCAAGCGACAG***TTTAAAGTGCTGCTTGCCATTCTG**-3′
11660-11704	5′-*GGCTGAGCAGCAGACTCAA***CAGGACAAAAATGCGCAGCAGCAGAG**-3′
8704-8749 (2 clones)	5′-*AGAGCTGGACAGCGATAC***CTGGCAGGCGGAGCTGCATATCGAAGTT**-3′
985-1032 (3 clones)	5′-*ACTTTATAGAGCATAAGCAGCGCAACACC***CTTATCTGGTTGCCGACGG**-3′
11651-11697	5′-*CCGAAAGAAGGCTGAGCAGCAGACTCAA***CAGGACAAAAATGCGCAGC**-3′
38757-38804	5′-*TCTCGATGATGGTTACGCCAG***ACTATCAAATATGCTGCTTGAGGCTTA-**3′
5182-5229	5′-*CCTGTCCGTGCAGGTTGTG***CTGGATACCGAGGCTGCAGTGTACAGCGG-**3′
11304-11352	5′-*TATGTTGCTCAGTTGCAGCGTTCCGGC***GATGAAGCCGGGGCATTGCAGG-**3′
15605-15653	5′-*CGATTGAAGGTCCGGTGGATGG***CTTAAAAAGCGTGCTGCTGAACAGTAC**-3′
40860-40909	5′-*AAGTGTGTGTTTACCGCAGCATTAAAGC***AGCAGGATGTTGTTCCTAACCT**-3′
40870-40919	5′-*TTACCGCAGCATTAAAGCAGCAGGATGTT***GTTCCTAACCTTGCCGGGAAT-**3′
1988-2036	5′-*GATTATTATGGGCCGCCACGA***CGATGAACAGACGCTGCTGCGTGTGGAT-**3′
Shotgun pACYC184	
1079-1128	5′-*CTGTATGCACGAACCCCCCCG***TTCAGTCCGACCGCTGCGCCTTATCCGGTA**-3′

Base numbering refers to the conventional orientation of λ genome; bold text, a terminal portion of GeoICI-cut DNA fragments, T4 DNA polymerase/dNTPs blunted and cloned into pUC19/SmaI; italic text, uncloned DNA sequence adjacent to cloned GeoICI-derived restriction fragment; grey box + horizontal arrow – GeoICI cognate site; vertical arrows – GeoICI cleavage positions.

complementary to the methylation catalytic motif whereas the reverse primer was complementary to the S-adenosylmethionine (cofactor – methyl group donor) binding motif of the *tspGWIRM* gene (Zylicz-Stachula *et al.* 2009). PCR results indicated the presence of the diagnostic segment within *Geobacillus* sp. genome. It

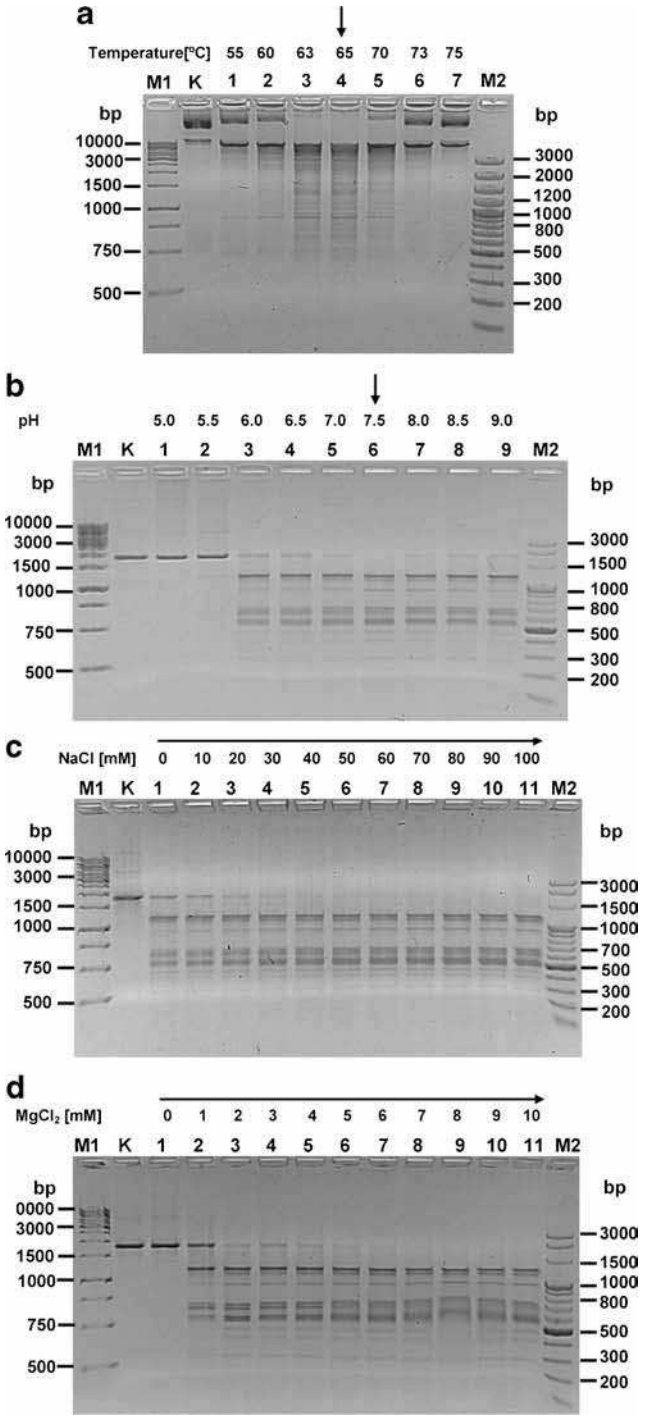


Figure 7. Evaluation of temperature, pH and salt concentration effect on GeoICI REase activity. (**a**) The optimum temperature range of GeoICI REase. 0.3 µg of pThermus DNA was digested with GeoICI in medium buffer for 1 h in the temperature range from 55 to 75°C. Lane M1 – GeneRuler™ 1 kb DNA Ladder (Thermo Fisher Scientific (Fermentas); selected bands marked); lane K – undigested pThermus DNA; lane 1 – 55°C; lane 2 – 60°C; lane 3 – 63°C; lane 4 – 65°C; lane 5 – 70°C; lane 6 – 73°C, lane 7 – 75°C, lane M2 – GeneRuler™ 100 bp Plus DNA Ladder (Fermentas), selected bands marked; arrow indicates the optimal temperature reaction using long DNA substrate. (**b**) The pH activity range of GeoICI REase. 0.3 µg PCR substrate (1789 bp) was digested with GeoICI in the pH range from 5.0 to 9.0 for 1 h at 50°C. Lane M1 – GeneRuler™ 1 kb DNA Ladder; lane K - undigested PCR fragment as described at Methods; lane1 – 5.0; lane 2 – 5.5; lane 3 – 6.0; lane 4 – 6.5; lane 5 – 7.0; lane 6 – 7.5; lane 7 – 8.0; lane 8 – 8.5; lane 9 – 9.0; lane M2 – GeneRuler™ 100 bp Plus DNA Ladder (selected bands marked); arrow indicates the optimal pH conditions. (**c**) The influence of ionic strength on GeoICI REase activity. 0.3 µg PCR substrate (1789 bp) was digested with GeoICI in the NaCl concentration range from 0 to 100 mM for 1 h at 50°C in the optimized reaction buffer. Lane M1 – GeneRuler™ 1 kb DNA Ladder (selected bands marked); lane K– undigested PCR substrate; lane1 – 0 mM; lane 2 – 10 mM; lane 3 – 20 mM; lane 4 – 30 mM; lane 5 – 40 mM; lane 6 – 50 mM; lane 7 – 60 mM; lane 8 – 70 mM; lane 9 – 80 mM; lane 10 – 90 mM; lane 11 – 100 mM; lane M2 – GeneRuler™ 100 bp Plus DNA Ladder (selected bands marked); arrow shows the increasing the salt concentration. (**d**) The influence of ionic strength on GeoICI REase activity. 0.3 µg PCR substrate (1789 bp) was digested with GeoICI in the $MgCl_2$ concentration range from 0 to 10 mM for 1 h at 50°C in the optimized reaction buffer. Lane M1 – GeneRuler™ 1 kb DNA Ladder (selected bands marked); lane K – undigested PCR substrate; lane 1 – 0 mM; lane 2 – 1 mM; lane 3 – 2 mM; lane 4 – 3 mM; lane 5 – 4 mM; lane 6 – 5 mM; lane 7 – 6 mM; lane 8 – 7 mM; lane 9 – 8 mM; lane 10 – 9 mM; lane 11 – 10 mM; lane M2 – GeneRuler™ 100 bp Plus DNA Ladder (selected bands marked); arrow shows the increasing the salt concentration.

remains to be determined whether a complete and functional *Thermus* sp. family enzyme gene is present.

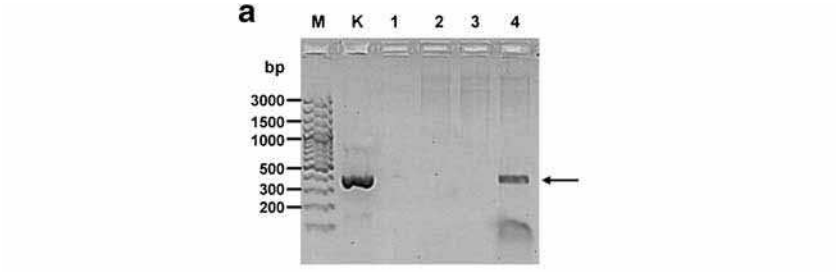
Isolation and purification of GeoICI was conducted in five steps: (i) crude lysate was cleared from nucleic acids and partially from acidic proteins with 0.5% poliethyleneimine; (ii) 35–65% ammonium sulphate salting-out cut; (iii) DEAE-cellulose chromatography, which resulted in the highest REase activity in a fraction eluted at a NaCl concentration of around 100 mM; (iv) Phosphocellulose P11 chromatography, where GeoICI was eluted at 800 mM NaCl and most contaminating non-specific nucleases were removed; (v) affinity purification on heparin-agarose, where GeoICI was eluted at 300 mM NaCl. Though no single protein band on SDS/PAGE was obtained, the GeoICI preparation was functionally homogeneous allowing for subsequent DNA enzymatic manipulations. Moreover, the described series of chromatographic steps allowed the correlation of fractions containing a particular band (65 kDa), which displayed maximum intensity on SDS-PAGE, with GeoICI activity profiles (not shown). The estimated GeoICI size corresponds to 66 kDa (figure 3), a value close to 61.7 kDa obtained for the mesophilic (30°C optimal reaction temperature) prototype enzyme, BbvI (Hattman *et al.* 1978). This suggests potential for the existence of homology between these enzymes, thus providing potential models for studying the relationship between the enzyme's amino acid sequence and its thermostability.

3.2 *Determination of GeoICI recognition and cleavage sites*

Initial digestions of standard substrates were not fully conclusive, because they apparently resulted in a mixture of complete

and partial digestion patterns, despite repeated digestion/precipitation procedures, as well as using an excess of the enzyme (figure 4). This happens for many Type IIS REases which often do not cleave substrate DNA completely.

Nevertheless, GeoICI cleavage patterns indicated a high frequency of cleavage, characteristic for a putative '4–5-bp' REase cognate site and similar to BbvI-characteristic cleavage patterns (figure 5a). Thus, further evaluation was conducted using short custom PCR substrates, with asymmetrically located, isolated 5′-GCAGC-3′ sequences. Digestions of four PCR substrates (see the Materials and Methods section) by GeoICI resulted in obtaining two distinct DNA bands (in addition to the uncut substrate) in each digest, visualized and analysed through 10% PAGE: 246 bp and 131 bp from 377 bp substrate, 140 and 64 from 204, 114 and 84 from 198, 96 and 103 from 199, which under the assumption that the cleavage site is further downstream of the recognition site, collaborates with the expected lengths, shown within the DNA sequences in figure 5a and b. As no additional bands were obtained, the possibility of existence of degenerated variants, which could not be differentiated in the initial large substrates digestions analysis, was eliminated. These include: 5′-GCAGC-3′, 5′-GCAGG-3′, 5′-CCAGG-3′, 5′-GCAGA-3′, 5′-GCAGT-3′. However, as agarose gel size estimation error was approximately 5%, the final determination with 1-bp precision came from run-off sequencing of the following GeoICI-digested DNA substrates: pUCmini, the 1789-bp and 198-bp PCR products (figure 6). The use of both forward and reverse primers determined cleavage points in both DNA strands as 8/12 nt downstream of the recognition site. Final validation of the recognition/cleavage sites came from shotgun cloning of λ DNA and pACYC184 digested with GeoICI. Insert-pUC19 vector junctions were sequenced and aligned through BLASTn online software and are summarized in Table 1. Further software prediction of cleavage frequency indicated that GeoICI is a 'high-frequency cutter' with 12 cognate sites within pUC19, 6 within pUCmini, 21 within pBR322, 15 within pACYC184, 200 within λ DNA and 117 within T7 bacteriophage DNA. Currently there are 8 isoschizomers of BbvI with determined DNA cleavage positions, 4 isoschizomers with cleavage points unknown, 1 putative isoschisomer and 2 neoisoschisomers (Roberts *et al.* 2015: REBASE, 16.08.2015).



b

Score	Expect	Identities	Gaps	Strand	Frame
405 bits(219)	8e-118()	247/274(90%)	0/274(0%)	Plus/Plus	

Features:

```
Query  76   GAGGGGCGGTTTGGCCCCGGGATGCGATCGCACTACGCCACGGAGGTGGCCCATCGGCTC  135
Sbjct  1    GAGGGGCGGTTTGGCCCCGGNATGCNATCGCACTACNCCACGGAGGTGGCCCATCGGNTC  60
Query  136  TACGCCTTGGAGGTCATGGTGGGCCCCTACGCCGTGGCTCAGCTCCGGCTCTCCCAGGCC  195
Sbjct  61   TACNCCTTGNAGGTCANGGNGGGCCCCTNCNCCGNGGCTCANCTCCGGCTCTCCCNGGCC  120
Query  196  ATCGAGGCTGAAGGTGGGCGGCTTCCCCCGGGGGGCCTCCAAGTCTACCTGGCGGACACC  255
Sbjct  121  ATCGAGGCTGANGGKGGGCGGNTTCCCCNGGGGGGCCTCCAAGTCTACCNGGCGGACNCC  180
Query  256  CTGGAGGCCCCCCGACGCCCCGCCCCTGGAGCGGGAGTTCTTCTACGAGCGCCTGGCGCAG  315
Sbjct  181  CCGGAGGCCCCCNACNCCCCNCCCCTGGAGCGGGANTTCTTCTACNAGCGCCTGGCGCAG  240
Query  316  GAGCGCCGGGAGGCGGCCCGGGTGAAGCGGGAGG  349
Sbjct  241  GANCNCCGGGAGGCGGCCCGGGTGAAGCGGGAGG  274
```

Figure 8. Evaluation of the presence of putative *Thermus* sp. enzymes family genes within *Geobacillus* sp genome as based DNA sequences analysis of the methylation catalytic and S-adenosylmethionine-binding motifs. (**a**) PCR amplification products analysis: Lane M**,** GeneRuler™ 100 bp Plus DNA Ladder, selected bands marked; lane K, PCR reaction with *Thermus* sp. GW genomic DNA; lane 1, PCR reaction without *Geobacillus* sp. genomic DNA added; lane 2, PCR reaction with *Geobacillus* sp. genomic DNA and forward primer; lane 3, PCR reaction with *Geobacillus* sp. genomic DNA and reverse primer; lane 4, PCR reaction with *Geobacillus* sp. genomic DNA and both primers added. (**b**) Comparison of the TspGWI MTase domain segment sequence with sequence obtained from PCR reaction using *Geobacillus* sp genomic DNA.

3.3 *Characterization of the GeoICI reaction requirements*

Initial GeoICI reactions indicated the 'medium salt' buffer conditions were nearly optimal, thus this buffer was used as a starting point for further biochemical investigations and DNA recognition/cleavage characterization. As GeoICI is a thermostable enzyme, the optimum temperature for DNA cleavage was determined on a substrate with a high GC content, i.e. a 8705 bp long derivative of a pET series vector with a segment cloned from *Thermus aquaticus*. As shown in figure 7, GeoICI activity is observed between 55°C and 73°C with a maximum at 63–65°C, which matches the optimal bacteria growth conditions. The activity temperatures of BbvI isoschizomers found thus far range between 30°C and 60°C, making GeoICI the most thermostable variant (Roberts *et al.* 2015: REBASE, 16.08.2015). As we have previously shown, REases isolated from thermophiles can be active in temperatures up to app. 15°C divergent both ways from the optimal growth temperature of the host bacterium. This may have different explanations, from playing a biological role in the regulation of cells' growth rate to an evolutionary background, such as a recent horizontal gene transfer from either a mesophile or thermophile with a different growth temperature range (Jezewska-Frackowiak *et al.* 2015). GeoICI restriction activity was observed in the pH range of 6.0-9.0, with a maximum at pH 7.0-8.5, thus pH 7.5 was used to determine the optimal ionic strength. GeoICI is fairly insensitive to a varying concentration of NaCl, with high activity observed in a range of 0-100 mM NaCl, whereas the nearly maximum activity was observed between 20 and 100 mM (figure 7). Like most Type II REases, GeoICI is not active without Mg^{2+} ions, which saturate the reaction from at a concentration of 2 mM (figure 7).

3.4 *Evaluation of the presence of genes coding for the family of bifunctional Thermus sp. enzymes in Geobacillus sp.*

We have conducted the second goal of this work by screening *Geobacillus* sp.' genomic DNA for putative genes coding for the Type IIS/IIG/IIC family of atypical, bifunctional *Thermus* sp. REases-MTases (Skowron *et al.* 2003; Skowron *et al.* 2013). Because these genes are long, i.e. over 3 kb, with a variable DNA sequence, the diagnostic marker chosen for amplification was a 450-bp segment, located within a conserved MTase domain, which includes S-adenosylmethionine-binding and methylation catalytic motifs. Analysis of the obtained PCR product (figure 8a, lane 4) revealed the presence of a consistent sequence found in the *Thermus sp*. GW region, coding for characteristic motifs of the TspGWI MTase domain. This also suggests that *Geobacillus* sp. possesses multiple R-M systems, thus we have temporarily assigned the new putative enzyme as GeoICI. This genome segment is currently being investigated and will be published elsewhere.

In conclusion, an enzyme useful for DNA manipulation biotechnologies was isolated and characterized, where both significant activity at high temperatures and at a broad range of salt concentration are advantageous over the mesophilic and, in our experience, unstable prototype, BbvI.

Acknowledgements

This work was supported by The Polish Ministry of Science and Higher Education grant DS/530-8640-D509-14 and BMN 538-8645-B743-15/14 (University of Gdansk, Faculty of Chemistry, Department of Molecular Biotechnology). Agnieszka Ozog is appreciated for the manuscript editing.

Author contributions: JZ screened environmental samples, conducted and interpreted the majority of experiments, co-drafted the manuscript and prepared figure drafts, OZ participated in experiment interpretations, MS provided environmental samples, participated in experiments and co-drafted the manuscript, AZS participated in experiment interpretations and prepared the final figures, JJF participated in experiment interpretations and co-drafted the manuscript, PMS conceived and coordinated the project, designed experiments, participated in experimental data analyses and co-drafted the manuscript. All authors have read and approved the final manuscript.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** 3389–3402

ATCC (*http://www.lgcstandards-atcc.org/?geo_country=pl*)

Azarko J and Wendt U 2011a Identification of microorganisms – comparison of biochemical and mass spectrometry method. *J Lab Diagn.* **47** 4

Azarko J and Wendt U 2011b Mass spectrometry – new method of identification of microorganisms. *Nowa Klinika, Medycyna Zakażeń.* **18** 4

Cuebas M, Sannino D and Bini E 2011 Isolation and characterization of arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils. *J. Basic Microbiol.* **51** 364–371

Edgar RC 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32** 1792–1797

Fox A 2006 Mass spectrometry for species or strain identification after culture or without culture: past, present, future. *J. Clin. Microbiol.* **44** 2677–2680

Gingeras TR, Milazzo JP and Roberts RJ 1978 A computer assisted method for the determination of restriction enzyme recognition sites. *Nucleic Acids Res.* **5** 4105–4127

Gingrich JC, Boehrer DM and Basu SB 1996 Partial CviJI digestion as an alternative approach to generate cosmid sublibraries for large-scale sequencing projects. *Biotechniques.* **21** 99–104

Green MR and Sambrook J 2012 *Molecular cloning: a laboratory manual* 4th ed. (New York: Cold Spring Harbor Laboratory Press)

Gupta R, Xu SY, Sharma P and Capalash N 2012 Characterization of MspNI (G/GWCC) and MspNII (R/GATCY), novel thermostable Type II restriction endonucleases from *Meiothermus* sp., isoschizomers of AvaII and BstYI. *Mol. Biol. Rep.* **39** 5607–56014

Hattman S, Keister T and Gottehrer A 1978 Sequence specificity of DNA methylases from *Bacillus amyloliquefaciens* and *Bacillus brevis*. *J. Mol. Biol.* **124** 701–711

Jezewska-Frackowiak J, Lubys A, Vitkute J, Zakareviciene L, Zebrowska J, Krefft D, Skowron MA, Zylicz-Stachula A, Skowron PM. 2015 A new prototype IIS/IIC/IIG endonuclease-methyltransferase TsoI from the thermophile *Thermus scotoductus*, recognising 5′-TARCCA(N11/9)-3′ sequences. *J. Biotechnol.* **194** 19–26

Krefft D, Zylicz-Stachula A, Mulkiewicz E, Papkov A, Jezewska-Frackowiak J and Skowron PM 2015 Two-stage gene assembly/cloning of a member of the TspDTI subfamily of bifunctional restriction endonucleases, TthHB27I. *J. Biotechnol.* **194** 67–80

Mead D, Swaminathan N, Van Etten J, Skowron PM 1995 Recombinant CviJI restriction endonuclease. (United States Patent Office no US005472872A)

Roberts RJ, Vincze T, Posfai J, Macelis D 2015 REBASE- a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43** (*http://rebase.neb.com*)

Skowron PM, Majewski J, Zylicz-Stachula A, Rutkowska SM, Jaworowska I and Harasimowicz-Slowinska RI 2003 A new *Thermus* sp. class-IIS enzyme sub-family: isolation of a 'twin' endonuclease TspDTI with a novel specificity 5′-ATGAA(N(11/9))-3′, related to TspGWI, TaqII and Tth111II. *Nucleic Acids Res.* **31** e74

Skowron PM, Swaminathan N, McMaster K, George D, Van Etten J and Mead DA 1995 Cloning and application of the two/three-base restriction endonuclease R.CviJI from IL-3A virus-infected *Chlorella Gene* **157** 37–41

Skowron PM, Vitkute J, Ramanauskaite D, Mitkaite G, Jezewska-Frackowiak J, Zebrowska J, Zylicz-Stachula A and Lubys A 2013 Three-stage biochemical selection: cloning of prototype class IIS/IIC/IIG restriction endonuclease-methyltransferase TsoI from the thermophile *Thermus scotoductus*. *BMC Mol. Biol.* **14** 17

Swaminathan N, McMaster K, Skowron PM and Mead DA 1998 Thermal cycle labeling: zeptomole detection sensitivity and microgram probe amplification using CviJI* restriction-generated oligonucleotides. *Anal. Biochem.* **255** 133–141

Vincze T, Posfai J and Roberts RJ 2003 NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* **31** 3688–3691

Xia YN, Burbank DE, Uher L, Rabussay D and Etten V 1987 JL: IL-3A virus infection of a *Chlorella*-like green alga induces a DNA restriction endonuclease with novel sequence specificity. *Nucleic Acids Res.* **15** 6075–6090

Yoon JH, Lee ST and Park YH 1998 Inter- and intraspecific phylogenetic analysis of the genus *Nocardioides* and related taxa based on 16S rDNA sequences. *Int. J. Syst. Bacteriol.* **48** 187–194

Zylicz-Stachula A, Harasimowicz-Slowinska RI, Sobolewski I and Skowron PM 2002 TspGWI, a thermophilic class-IIS restriction endonuclease from *Thermus* sp., recognizes novel asymmetric sequence 5′-ACGGA(N11/9)-3′. *Nucleic Acids Res.* **30**, e33

Zylicz-Stachula A, Bujnicki JM and Skowron PM 2009 Cloning and analysis of bifunctional DNA methyltransferase/nuclease TspGWI, the prototype of a *Thermus* sp. family. *BMC Mol. Biol.* **10** 52

Zylicz-Stachula A, Żołnierkiewicz O, Jeżewska-Frąckowiak J and Skowron PM 2011a Chemically-induced affinity star restriction specificity: a novel TspGWI/sinefungin endonuclease with theoretical 3-bp cleavage frequency. *Biotechniques* **50** 397–406

Zylicz-Stachula A, Żołnierkiewicz O, Śliwińska K, Jeżewska-Frąckowiak J and Skowron PM 2011b Bifunctional TaqII restriction endonuclease: redefining the prototype DNA recognition site and establishing the Fidelity Index for partial cleaving. *BMC Biochem.* **12** 62

Zylicz-Stachula A, Zolnierkiewicz O, Lubys A, Ramanauskaite D, Mitkaite G, Bujnicki JM and Skowron PM 2012 Related bifunctional restrictionendonuclease-methyltransferase triplets: TspDTI, Tth111II/TthHB27I and TsoI with distinct specificities. *BMC Mol. Biol.* **13** 13

Zylicz-Stachula A, Zolnierkiewicz O, Sliwinska K, Jezewska-Frackowiak J and Skowron PM 2014 Modified 'one amino acid-one codon' engineering of high GC content TaqII-coding gene from thermophilic *Thermus aquaticus* results in radical expression increase. *Microb. Cell Factories* **13** 7

MS received 20 September 2015; accepted 18 January 2016

Corresponding editor: B Jagadeeshwar Rao