

# ***TspGWI*, a thermophilic class-IIS restriction endonuclease from *Thermus* sp., recognizes novel asymmetric sequence 5'-ACGGA(N<sub>11/9</sub>)-3'**

**Agnieszka Żylicz-Stachula<sup>1</sup>, Renata I. Harasimowicz-Słowińska<sup>1</sup>, Ireneusz Sobolewski<sup>1</sup> and Piotr M. Skowron<sup>1,2,\*</sup>**

<sup>1</sup>EURx Ltd, ul. Jaśkowa Dolina 29, 80-286 Gdańsk, Poland and <sup>2</sup>Department of Microbiology, Technical University of Gdańsk, Gdańsk, Poland

Received November 27, 2001; Revised and Accepted February 9, 2002

## **ABSTRACT**

**A novel prototype class-IIS restriction endonuclease, *TspGWI*, was isolated from the thermophilic bacterium *Thermus* sp. GW. The recognition sequence and cleavage positions have been established: *TspGWI* recognizes the non-palindromic 5-bp sequence 5'-ACGGA-3' and cleaves the DNA 11 and 9 nt downstream in the top and bottom strand, respectively. In addition, an accompanying endonuclease, *TspGWII*, an isoschizomer of *PstI*, was found in *Thermus* sp. GW cells.**

## **RESULTS AND DISCUSSION**

The general characteristics of type II restriction endonucleases are that they require Mg<sup>2+</sup> ions as the only obligatory co-factor and recognize a palindromic 4–8-bp cognate site, where cleavage takes place at precisely defined positions within the recognition site (1). The type II enzymes are highly heterogeneous, with a distinct subclass, IIS, which differs from the class-II paradigm by its ability to recognize asymmetric sequences and cleave further downstream at defined distances of 1–20 nt (2). Currently there are 73 known prototypes of endonucleases that recognize asymmetric sites (1). However, only 43 of them meet the criteria of class-IIS. The remaining 30 enzymes either (i) cleave within asymmetric sites, (ii) have not yet had their cleavage positions determined or (iii) cleave on both sides outside their recognition sites (1).

*Thermus* sp. GW was cultivated aerobically at 60°C in a modified Luria broth (0.5% tryptose, 0.3% yeast extract, 0.2% NaCl, pH 7.2) supplemented with trace elements. Since class-IIS restriction endonuclease *TspGWI* activity is only present in minute quantities in the host strain, an extensive purification was required to obtain a sufficient amount of enzyme for further analysis. The purification procedure included (i) 0.4% polyethyleneimine removal of nucleic acids from the crude extract in the presence of 100 mM NaCl, (ii) ammonium sulfate 30–50% saturation cut, (iii) phosphocellulose chromatography, (iv) heparin–agarose chromatography, (v) Cibacron blue–agarose chromatography and (vi) DEAE–Sephadex chromatography. The purified preparation was free from non-specific nucleolytic

activities. In the course of the purification of *TspGWI* an accompanying restriction endonuclease, *TspGWII*, a thermophilic isoschizomer of *PstI*, was found. *TspGWII* recognizes 5'-CTG-CAG-3' and cleaves after the A residue, leaving 4 nt 3'-protruding ends (not shown).

The *TspGWI* recognition site was established using two procedures: (i) assessment of the digestion pattern on pUC19, pACYC184 and pBR322 DNAs and (ii) cloning and sequencing of *TspGWI* restriction fragments of pBR322 and bacteriophage lambda DNAs (Table 1). The analysis of the cleavage pattern and the mapping of *TspGWI* sites present in pUC19 (Fig. 1A and B) suggested that this recognition site is 5'-ACGGA-3'. There are two such sites in pUC19, four in pACYC184 and five in pBR322. Cleavage of pUC19 with *TspGWI* resulted in a complete digestion. However, cleavage of pBR322 and pACYC184 DNAs showed a stable partial digestion pattern, where all sites except one are efficiently cleaved (Fig. 1C–E). The refractory sites, cleaved at a substantially slower rate, were mapped and they turned out to be located within the tetracycline resistance gene, present in both plasmids. One of the possible explanations for the phenomenon of refractory cleavage might be that it is an effect of the immediate sequence segment surrounding the *TspGWI* site. The neighboring bases, which are present on both sides of the *TspGWI* recognition sequences in pBR322 and pACYC184, are shown in bold: 5'-AACGGAT-3'. The putative *TspGWI* cognate site was further investigated through the digestion of pBR322 DNA and lambda DNA with *TspGWI*, followed by repair of the *TspGWI* restriction fragment termini with T4 DNA polymerase/dNTPs (3) and by cloning the resulting restriction fragments into the *SmaI* site of a modified pTZ18u vector (4). The vector-insert junctions of the resulting clones were sequenced using the ABI Prism 310 automated sequencer with ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit and then analyzed using ABI Sequencing Analysis 3.4.1 software (Applied Biosystems, Foster City, CA) and Hitachi DNASIS 2.5 software (Hitachi Software Engineering Co., Yokohama, Japan). A total of 58 junction sequences were compared. Twenty-five junctions derived from bacteriophage lambda cloned restriction fragments are exemplified in Table 1, where eight junctions are for the top strand of the 5'-ACGGA-3' and 17 for the bottom strand of the 5'-TCCGT-3' reverse complement

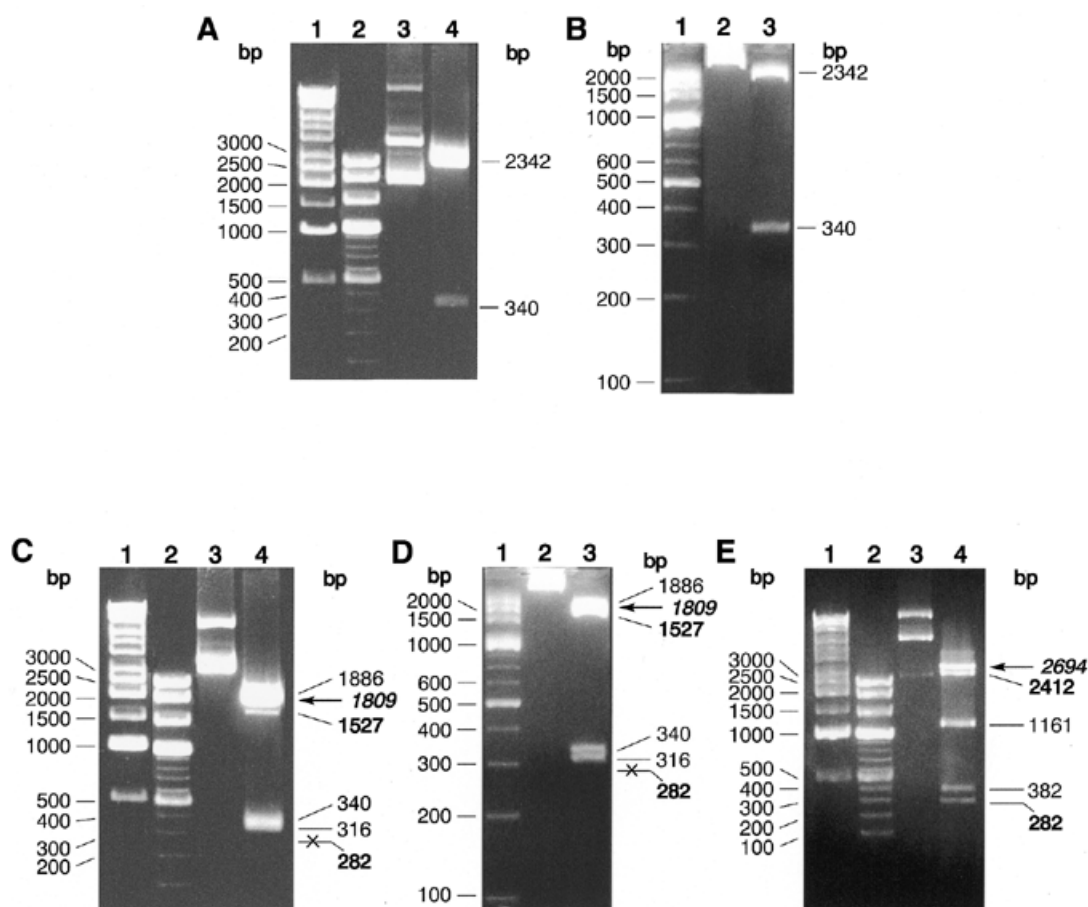
\*To whom correspondence should be addressed. Tel: +48 58 3452155; Fax: +48 58 3417423; Email: piotrs@eurx.com.pl

**Table 1.** Determination of the *Tsp*GW1 recognition sequence and cleavage positions by shot-gun cloning and sequencing of *Tsp*GW1 restriction fragments

Sequence position in lambda genome.	DNA sequence flanking <i>Tsp</i> GW1 cleavage site in bacteriophage lambda DNA.
152-185	5'-...TGTCGTTTCCTTCTCTGTTTTG <b>TCCGT</b> GGAAT...-3' ↑ ↓
1064-1097	5'-...CGTTGAGCCGACTATTCGTGATAT <b>TCCGT</b> CGCTG...-3' ↑ ↓
2189-2224	5'-...CGTCT <b>ACGGA</b> AAGCCGGTGGCCAGCATGCCACGTAA...-3' ↑ ↓
2189-2222 *	5'-...ACGTGGCATGCTGGCC <b>ACCGGCTT</b> <b>TCCGT</b> AGACG...-3' ↑ ↓
3869-3904	5'-...AGGAT <b>ACGGA</b> TAACGGCTACTCCGTGTTTGAGCAGT...-3' ↑ ↓
4292-4325	5'-...ATCAGGAAATTTT <b>GCCCAGCAGG</b> <b>TCCGT</b> GAAAC...-3' ↑ ↓
4615-4650	5'-...TGATG <b>ACGGA</b> CCACGACAGGCCCGCAGTTATCAGGT...-3' ↑ ↓
10779-10814 *	5'-...CAGCC <b>ACGGA</b> CTTTGCCCGCCTGCAAGCTGCGTGGC...-3' ↑ ↓
10781-10814	5'-...CACGCAGCTTGCAGG <b>CGGGCAAAG</b> <b>TCCGT</b> GGCTG...-3' ↑ ↓
12562-12597*	5'-...ACAGC <b>ACGGA</b> ACGGGTGAAGCTGCGCCAGTTCTGCT...-3' ↑ ↓
16454-16489	5'-...GCGGC <b>ACGGA</b> GCCGCGCATCACCTGTAATGCGTACC...-3' ↑ ↓
17205-17240 *	5'-...TTCTC <b>ACGGA</b> TACTCAGCAGCGGAACAGTCGCTGG...-3' ↑ ↓
19472-19505	5'-...GGTATACAGATTAAT <b>CCGGCAGCG</b> <b>TCCGT</b> CGTTG...-3' ↑ ↓
19546-19579 *	5'-...TTTATAACCGACCCCA <b>ACGATGAA</b> <b>TCCGT</b> CAGTA...-3' ↑ ↓
19924-19958	5'-...CCATG <b>ACGGA</b> GGATGATGCCCGGCCGGAGGTGCTG...-3' ↑ ↓
19977-20010	5'-...GTGGAAGAGGTGGCG <b>CGTAACGCG</b> <b>TCCGT</b> GGTGG...-3' ↑ ↓
19975-20010 *	5'-...CCACC <b>ACGGA</b> CGCGTTACGCCACCTCTCCACCA...-3' ↑ ↓
21994-22029	5'-...CAACC <b>ACGGA</b> CCATAAAAATTTATAATCTGCTGGCC...-3' ↑ ↓
22319-22352 *	5'-...CGGATCCGGAACAGT <b>TTTTCTGCT</b> <b>TCCGT</b> ATCCT...-3' ↑ ↓
30971-31004	5'-...AAGCCCGCTGCCAGAAAATGCAT <b>TCCGT</b> GGTTG...-3' ↑ ↓
31291-31324 *	5'-...GTAGCGCAATCACT <b>TTCTGCTAC</b> <b>TCCGT</b> TACAA...-3' ↑ ↓
32664-32631 *	5'-...CGTTGCCAACCAAGT <b>CGGCCTTAA</b> <b>TCCGT</b> GGACG...-3' ↑ ↓
33986-34019	5'-...CTGTTTAGTTACGAG <b>CGACATTGC</b> <b>TCCGT</b> GTATT...-3' ↑ ↓
34316-34283*	5'-...GTGATGGAACAATA <b>CCAGGACTA</b> <b>TCCGT</b> ATGAC...-3' ↑ ↓
43565-43598	5'-...TCTTGCCATAAAGC <b>AGATGAACT</b> <b>TCCGT</b> TAATC...-3' ↑ ↓
44551-44518*	5'-...TATCATGCCGTTAAT <b>ATGTTGCCA</b> <b>TCCGT</b> GGCAA...-3' ↑ ↓

Base numbering refers to the conventional orientation of the lambda genome. Bold, a terminal portion of *Tsp*GW1-cut restriction fragment, T4 DNA polymerase repaired and cloned into pTZ18u derivative; italic, uncloned bacteriophage lambda DNA sequence adjacent to cloned *Tsp*GW1-derived restriction fragment; box with horizontal arrow, *Tsp*GW1 recognition sequence; vertical arrows, *Tsp*GW1 cleavage positions.

\*Sequence obtained from *Tsp*GW1 restriction fragments cloned in reverse complement orientation.

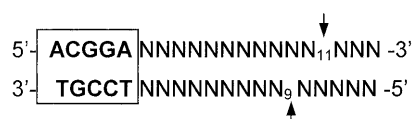


**Figure 1.** Digestion pattern and site preference of *TspGWI* on pUC19, pBR322 and pACYC184 plasmid DNAs. (A) *TspGWI* cleavage of pUC19 DNA, 1% agarose/TAE. Lane 1, 1 kb ladder; lane 2, 100 bp ladder; lane 3, untreated pUC19 DNA; lane 4, *TspGWI*-cut pUC19 DNA. (B) *TspGWI* cleavage of pUC19 DNA, 3.5% NuSieve GTG agarose/TAE. Lane 1, 100 bp ladder; lane 2, untreated pUC19 DNA; lane 3, *TspGWI*-cut pUC19 DNA. (C) cleavage of pBR322 DNA, 1% agarose/TAE. Lane 1, 1 kb ladder; lane 2, 100 bp ladder; lane 3, untreated pBR322 DNA; lane 4, *TspGWI*-cut pBR322 DNA. Out of five *TspGWI* sites in pBR322, one 5'-AACGGAT-3' variant is cleaved inefficiently, resulting in fragments of 1527 bp (faint band, marked in bold) and 282 bp (not visible on reproduced picture, marked in bold and crossed). The partial digestion band of 1809 is indicated by bold italics and a horizontal arrow. (D) *TspGWI* cleavage of pBR322 DNA, 3.5% NuSieve GTG agarose/TAE. Lane 1, 100 bp ladder; lane 2, untreated pBR322 DNA; lane 3, *TspGWI*-cut pBR322 DNA. (E) *TspGWI* cleavage of pACYC184 DNA, 1% agarose/TAE. Lane 1, 1 kb ladder; lane 2, 100 bp ladder; lane 3, untreated pACYC184 DNA; lane 4, *TspGWI*-cut pACYC184 DNA. Bands of 2412 and 282 bp, where intensity is decreased due to slow cleavage rate of *TspGWI* site variant 5'-AACGGAT-3', are marked in bold. Partial digestion band of 2694 bp is marked in bold italics and a horizontal arrow. Selected band sizes of marker DNAs are shown on the left of each panel.

orientation. Comparison of the junction sequences confirmed the presence of the 5'-ACGGA-3' site at either defined distances near the ends of a cloned restriction fragment or in pBR322/lambda DNA regions which, prior to *TspGWI* cleavage, were continuous with their corresponding cloned fragment. The sequences presented in Table 1 were selected to show all found combinations of the neighboring 1 bp, flanking both sides of the *TspGWI* recognition sequences. The computer prediction revealed 107 5'-ACGGA-3' sites in lambda. Four out of the 16 possible combinations of 1-bp neighborhoods are under-represented in the lambda genome and they were also not detected during the junction sequence analysis (marked in bold): 5'-AACGGAT-3' (5/107), 5'-AACGGAC-3' (6/107), 5'-TACGGAC-3' (1/107) and 5'-TACGGAG-3' (0/107). Since the 5'-AACGGAT-3' site present in the tetracycline resistance gene is slowly cleaved this may apply to lambda as well. This would decrease the cloning efficiency of 5'-AACGGAT-3' fragments. Whether the absence of 5'-AACGGAC-3' and 5'-

TACGGAC-3' sequences amongst the analyzed junctions is also caused by diminished reaction efficiency remains to be determined.

The *TspGWI* recognition sites from the sequenced junctions appear in four possible configurations: as top or bottom strand in either an insert or an uncloned segment of adjacent pBR322/lambda DNA regions (Table 1). The comparison of various clone configurations allowed the deduction of the cleavage positions, which are located further downstream from the 5'-ACGGA-3' sequence, at the 11 and the 9 nucleotide in the top and bottom strands (vertical arrows), respectively, leaving 2 nt 3'-overhangs:



The recognition sequence and cleavage points of *Tsp*GW1 show similarities to those of the two known *Thermus*-derived restriction endonucleases of type-IIS: *Taq*II from *Thermus aquaticus*, GACCGA(N<sub>11/9</sub>)-3' or CACCCA(N<sub>11/9</sub>)-3' (5), and *Tth*111III from *Thermus thermophilus* 111, CAARCA(N<sub>11/9</sub>)-3' (6). The conservation of cleavage positions can be observed, as well as of the first and the last A of the *Tsp*GW1 recognition site. This may indicate a common evolutionary origin for all these endonucleases.

The optimum reaction conditions for *Tsp*GW1 are 50 mM Tris-HCl pH 8.5 at 25°C, 10 mM MgCl<sub>2</sub>, 10 mM DTT and 3 mM spermidine. The enzyme is active between 42 and 85°C, with a maximum at 65 and 75°C. The enzyme can be inactivated by 20 min incubation at 89°C.

#### ACKNOWLEDGEMENTS

The helpful suggestions regarding the manuscript by Prof. Józef Kur and Jakub Zaleski as well as figure preparation by Barbara Skowron-Zaleska are appreciated. The help in

isolation of bacteria by Izabela Jaworowska and Robert Atras are acknowledged. This project was supported by EURx Ltd funds available to P.M.S.

#### REFERENCES

1. Roberts,R.J. and Macelis,D. (2001) REBASE—restriction enzymes and methylases. *Nucleic Acids Res.*, **29**, 268–269.
2. Szybalski,W., Kim,S.C., Hasan,N. and Podhajaska,A.J. (1991) Class-IIS restriction enzymes—a review. *Gene*, **100**, 13–26.
3. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
4. Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Eng.*, **1**, 67–74.
5. Barker,D., Hoff,M., Oliphant,A. and White,R. (1984) A second type II restriction endonuclease from *Thermus aquaticus* with unusual sequence specificity. *Nucleic Acids Res.*, **12**, 5567–5581.
6. Shinomiya,T., Kobayashi,M. and Sato,S. (1980) A second site-specific endonuclease from *Thermus thermophilus* 111, Tth111III. *Nucleic Acids Res.*, **8**, 3275–3285.