Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 363 (2005) 211–220
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Housekeeping *recA* gene interrupted by group II intron in the thermophilic *Geobacillus kaustophilus*

Gab-Joo Chee *, Hideto Takami

Microbial Genome Research Group, XBR, Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima Yokosuka 237-0061, Japan

Received 9 May 2005; received in revised form 11 August 2005; accepted 12 August 2005
Available online 19 October 2005
Received by T. Ikemura

Abstract

Most of group II introns are found in intergenes and CDSs with unknown functions, but not in housekeeping genes. In particular, no group II intron within the housekeeping *recA* gene has been reported either in eukaryotic genomes or in prokaryotic genomes. In this study, we found that the *recA* gene of the thermophilic *Geobacillus kaustophilus* genome is interrupted by a group II intron (Gk. Int1), and that Gk.Int1 can splice in temperatures above 70 °C in vivo. Here, we report the first prokaryotic group II intron to be found in a housekeeping gene, the characteristics of its self-splicing in vivo and in vitro, and our conclusion that the *recA* gene functions through the self-splicing of Gk.Int1. It is suggested that the amelioration of Gk.Int1 intron has occurred recently, and that it is still in the process of evolution to the recipient genome.

Keywords: Group II intron; Housekeeping gene; *recA* gene; *Geobacillus kaustophilus*; Self-splicing

1. Introduction

Group II introns are retroelements and large autocatalytic RNAs (Peebles et al., 1986). Group II introns are widely believed to be the evolutionary progenitors of eukaryotic spliceosome introns due to their splicing mechanism (Lehmann and Schmidt, 2003). With increasing genomic sequence data, more and more group II introns and ORF-less introns are being reported in various bacterial and archaeal genomes (Dai and Zimmerly, 2003; Toro, 2003; Robart et al., 2004). Organella group II introns are substantially inserted into essential genes, whereas prokaryotic group II introns are hardly found in housekeeping genes. If anything, prokaryotic group II introns are associated with mobile elements or plasmids, in intergenic region, and some interrupt ORFs of unknown function (Martinez-Abarca and Toro, 2000; Dai and Zimmerly, 2002). On the other hand, it has been reported that the *groEL* gene encoding a heat-shock protein (Hsp60) was affected by the insertion of a group II intron in the genome of *Azotobacter vinelandii* (Adamidi et al., 2003). However, the group II intron was only inserted in the region of the stop codon, and eventually no change occurred in the amino acid sequence of the GroEL protein, although another sequence for the stop codon was created by the self-splicing.

RecA protein is ubiquitous and the most conserved proteins in prokaryotic organisms (Karlin and Brocchieri, 1996). It is well known as a multifunctional protein that involve in homologous recombination, DNA repair, and the SOS response. In *B. subtilis*, the transcription of the *recA* gene is negatively regulated by the DinR protein (named LexA in *E. coli*), which binds to the upstream regions of the SOS genes containing *recA* gene (Winterling et al., 1997).

Geobacillus kaustophilus HTA426 (JCM12893) is a thermophilic *Bacillus*-related species isolated from the deepest point of the Mariana Trench (11,000 m) (Takami et al., 1997). *G. kaustophilus* is a thermophilic bacterium growing up to 74 °C (Takami et al., 2004a). The complete genomic sequence of the strain HTA426 has been determined by a whole-genome shotgun sequencing approach (Takami et al., 2004b). We discovered the *recA* gene interrupted by putative group II intron in the strain

Abbreviations: CDS, protein coding sequence; EBS, exon binding site; IBS, intron binding site; ORF, open reading frame; RSCU, relative synonymous codon usage.

* Corresponding author. Tel.: +81 46 867 9651; fax: +81 46 867 9645.
E-mail address: cheegj@jamstec.go.jp (G.-J. Chee).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.08.017

HTA426 genome. We describe that a putative group II intron is self-splicing in vivo and in vitro, and that the *recA* gene codes for a functional RecA protein in *E. coli*. Furthermore, we discuss the amelioration of group II intron to the *G. kaustophilus* genome.

2. Materials and methods

2.1. Preparation of DNA and RNA

G. kaustophilus HTA426 was grown in Luria-Bertani (LB) medium at 60 °C. The *E. coli* strains used were DH5α and C600. When required, ampicillin (100 μg/ml) or kanamycin (25 μg/ml) was added to the medium as an antibiotic. The chromosomal DNA of *G. kaustophilus* was isolated with a QIAGEN Genomic DNA Purification Kit in accordance with the manufacturer's protocol (Qiagen). Total RNA was purified from cells grown in the LB medium to an optical density of 1.0 at 600 nm, as described previously (Igo and Losick, 1986) but with some modifications by washing with ice-cold 10 mM Tris buffer (pH 7.4) before using LETS buffer and by purification of the RNA pellet with the RNeasy kit (Qiagen).

2.2. PCR and DNA sequencing

All PCRs were carried out on a GeneAmp PCR System 9700 (Perkin-Elmer). The DNA sequence was determined using the dRhodamine (dichlororhodamine) Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). The reaction products were analyzed by an ABI PRISM 377 DNA Sequencer. The primers used for sequencing the inserted fragments were the following: DAMAS5, 5′-GTTTGCCG-AGTCTGGATTGATC-3′; DAMAS6, 5′-CGCCATTACAA-AGGAATGTCTGG-3′; DAMAS7, 5′-GGGTACAATTGGGT-GATAGAGG-3′; DAMAS8, 5′-GCAGGGGGAATGCCGAA AAATG-3′; DAMAS9, 5′-ACCATTGTGCTCTTACGGTC-AG-3′; DAMAS10, 5′-GGTAATAGTCACCCTCGGCATT-3′.

2.3. Plasmid construction

In order to obtain the entire *recA* ORF, chromosomal DNA was used as a template for PCR with oligonucleotide primers DAMAS1, 5′-CCCAAGCTTGCTTGGCAATGGCC GCAAAAC-3′, and DAMAS2, 5′-CGGGGTACCAGGCTT-ATCGCTCTCCTTGGCA-3′, from 64 bp upstream of the start codon to 76 bp downstream of the stop codon. The PCR-generated fragment was cloned into the *Hin*dIII–*Kpn*I site of pUC19 and the *Not*I–*Kpn*I site of pSD64TF, in order to generate pUC*recA*[int$^+$] and pSD*recA*[int$^+$], respectively. The *recA* gene of pSD*recA*[int$^+$] was under the control of the SP6 promoter. A pUC*recA*[int$^-$] plasmid was constructed by cloning the intronless version of the *recA* gene yielded via reverse transcription-PCR (RT-PCR), using the primers above. The underlined sequences indicate the introduced restriction sites. In DAMAS1, the restriction site for pSD*recA*[int$^+$] was introduced into *Not*I instead of *Hin*dIII.

2.4. In vivo and in vitro splicing

The pSD*recA*[int$^+$] plasmid was linearized with *Hin*dIII for in vitro synthesis of RNA. The transcription reaction was done with 0.5 mg of plasmid template and SP6 RNA polymerase (Ambion), in accordance with the manufacturer's protocol. cDNA synthesis of the ligated exon products was performed using 5 U AMV Reverse Transcriptase XL (Takara) at 50 °C for 30 min, as per the manufacturer's protocol, with 1 mg of total RNA or RNA from in vitro synthesis as a template, 5 mM $MgCl_2$ and the primers specific to *recA*-A and *recA*-B (DAMAS3, 5′-ATCGTCCGGGAAAACGACCGTTG-3′; DAMAS4, 5′-GG-ACGGAAGCGTAAAACTTGAGC-3′). The cycling conditions were as follows: 1 cycle at 94 °C for 2 min; 40 cycles at 94 °C for 30 s, 57 °C for 30 s, and 72 °C for 1 min/kb.

2.5. DNA damage survival assays

G. kaustophilus and *E. coli* strains (C600, and DH5α containing pUC19, pUC*recA*[int$^+$], and pUC*recA*[int$^-$] plasmids) were grown at 55 and 37 °C in an LB liquid medium up to the exponential stage (OD_{600}, 0.9 to 1.0), respectively. The cells were serially diluted in 10% LB medium and plated on LB agar plates. The plates were immediately irradiated with various doses of UV (254 nm) and incubated overnight in the dark at the temperatures stated above. The colonies were counted in order to determine the rates of survival. The total number of viable cells from among the serially diluted unirradiated cells was counted.

2.6. Investigation of amelioration

Amino acid alignments were performed by use of the multiple-alignment program Clustal W (Thompson et al., 1994). Phylogenetic analysis was carried out by (Saitou and Nei, 1987) the neighbor-joining method using the DNADIST and NEIGHBOR program in the PHYLIP package (version 3.57) (Felsenstein, 1995). The sequence data for the prokaryotic genomes were taken from the GenBank FTP site (ftp://ftp.ncbi.nih.gov/genomes/Bacteria). The codon usage numbers were converted into relative synonymous codon usage (RSCU) (Sharp and Li, 1986) values. Putative group II intron ORFs (reverse transcriptase) were identified by Blast searches at the NCBI web site (http://www.ncbi.nlm.nih.gov/BLAST) using a selection of known group II intron ORFs as query sequences. Next, the regions of upstream and downstream from each of the identified reverse transcriptase genes were searched for the consensus sequences (GUGYG-AY) at the 5′ and 3′ splicing sites of group II introns.

3. Results and discussion

3.1. Putative recA gene containing an intron

The genome sequence of *G. kaustophilus* HTA426 as a whole revealed that its *recA* gene (GK1295, GenBank accession nos. BA000043) was interrupted by a group II intron, which is a 2782